PAPER

Check for updates

Using social influence models to characterize student interest in a general chemistry peer-led team learning setting

Jacob D. McAlpin, [ID] Ushiri Kulatunga [ID] and Jennifer E. Lewis [ID] *

Motivation helps drive students to success in general chemistry, and active learning environments with social interactions has consistently shown to improve motivation. However, analyzing student outcomes in an interactive environment is best done by considering students not as isolated units but as working together and influencing each other. Therefore, we used social network analysis with self-determination theory as a framework for understanding motivation and social comparison theory as a framework for understanding how students influence each other. When analyzing an undergraduate general chemistry course that has incorporated peer-led team learning using data from the Learning Climate Questionnaire and Intrinsic Motivation Inventory, a series of progressively sophisticated statistical models with data gathered from 270 students shows that perceived competence and relatedness predict student interest in the activities with their peer-led sessions. However, we also found evidence that students tend to become polarized in their interest toward peer-led team learning activities, which is one possible outcome of social comparisons with their peers. In addition to these findings, this project demonstrates how social network analysis can expand how chemistry education researchers consider relational data and the effects of non-independent data on statistical analysis.

Introduction

Many chemistry education researchers have investigated student motivation to help understand the role of motivation in course performance and have explored interventions to improve motivation as a way to promote student success (Flaherty, 2020). Such work has demonstrated a positive relationship between student motivation in chemistry and course performance in traditional classrooms (Ferrell *et al.*, 2016) and flipped classrooms (Hibbard *et al.*, 2016). Additionally, a study looking at the relationship between motivation and post-graduation plans from an international sample of chemistry students from Australia, New Zealand, and the UK found interest in chemistry to be the predominant factor behind the decision to choose to study chemistry (Ogunde *et al.*, 2017).

Considering the potential impacts of motivation on student success, chemistry education researchers have investigated how pedagogy affects motivation. One of the findings from this work is the recognition of a positive relationship on motivation gained from using active learning pedagogies such as the use of Study Periods and Discussion Groups (Cicuto and Torres, 2016), hands-on science teaching (Jurišević *et al.*, 2012), and Process Oriented

Guided Inquiry Learning (POGIL; Southam and Lewis, 2013). One particular active learning technique with promise to support student motivation in chemistry is Peer-Led Team Learning (PLTL; Gosser and Roth, 1998; Liu *et al.*, 2018). Under this pedagogical method, students in large lecture courses are placed into groups under the supervision of a peer leader who is an undergraduate student who has already successfully passed the course. As part of the PLTL environment, students have prolonged engagement with peers in their class over the length of the course. While this prolonged engagement has the potential to support student motivation and persistence, it also has implications for studying PLTL. As students have the ability to influence each other, researching the potential impacts of PLTL should acknowledge that students are not isolated units but are incorporated into groups (Stevens, 2007; Theobald, 2018).

Non-independent data and social networks

This concern is especially true in many statistical analyses where independence of data is fundamental, such as multiple regression or ANOVA. Even a small violation of independence can have a profound effect on the chance of making a Type I error (Stevens, 2007). The rate of a Type I error (α) represents the chance that the null hypothesis is rejected when it actually holds true, or said another way, α is the chance of a false positive. When examining the effect of intraclass correlation on

Department of Chemistry, University of South Florida, 4202 E. Fowler Avenue, Tampa, Florida 33620, USA. E-mail: jennifer@usf.edu

This journal is © The Royal Society of Chemistry 2023

Chem. Educ. Res. Pract., 2023, **24**, 1003–1024 | 1003

ANOVA Type I error rates with a nominal value of 0.05, Scariano and Davenport (1987) found the effect of violating independence to be quite profound. When comparing three groups of ten subjects per group with an intraclass correlation of 0.3 (moderate dependence), the actual α was 0.4917 compared to the nominal 0.05. They also found when holding intraclass correlation constant and increasing the number subjects per group, the gap between the observed α and the nominal α increases even more. This inflated chance of false positive results is important. When an active learning technique involves direct interactions among students, the assumption of independence may not be tenable, as students consistently have the opportunity to influence each other as part of the learning process (Theobald, 2018).

One method to analyze a population while considering how individuals might affect each other is through use of social influence models which have a few ways to model relational effects (Lane *et al.*, 2019; Leenders, 2002). This technique falls under the broader category of Social Network Analysis (SNA) which is a set of processes and tools largely developed within sociology and mathematics for considering relational data (Wasserman and Faust, 1994) and has been promoted as a valuable tool in discipline-based education research (Grunspan *et al.*, 2014). Social networks are maps of individuals (nodes) and the connections between them (edges/ties). The connections in a social network can either be directed, where a connection is not necessarily reciprocal (student A loans book to student B), or undirected, where the connection is reciprocal (student A and B study together). The methods of SNA were developed from graph theory and can allow education researchers to consider behaviors and social dynamics in the classroom.

The applications of SNA in education research are quite varied so we will just present a sampling here. First, Brewe *et al.* (2012) used SNA to investigate student communities in a physics learning center. They found that a student's centrality (term describing how well an individual is connected within a network) in the student community was predicted by days per week spent in physics learning center and whether or not the student was a physics major. However, a student's centrality in the network was not significantly predicted by gender and ethnicity variables suggesting an equitable learning environment. At the same location, Dou *et al.* (2016) further found that they could identify a correlation between a student's self-efficacy and the student's network centrality. Another option within SNA is to look more directly at how individual students affect each other. This was demonstrated by Vitale *et al.* (2016) when they reported for a group of graduate students in an Italian university. They found that while formal groups of students formed temporarily did not relate to graduate school performance, informal groups based on mutual interest and goals were predictive of performance. Within chemistry education research specifically, a similar set of tools to SNA has been used to analyze discussion in a POGIL style physical chemistry course (Liyanage *et al.*, 2021). Within that course, three distinct patterns of student engagement were observed that seemed related to instructor facilitation strategies. Additionally, SNA

was used to map the interactions of undergraduates, faculty, and graduate students at a virtual undergraduate poster session to show the different kinds of interactions between those groups (Bongers, 2022).

Student experiences entering chemistry

When considering the impacts of socialization and interaction on student performance and affective outcomes, we also recognize that students come into general chemistry with a variety of life experiences and history with chemistry. For example, transfer students come into a university setting after having experienced some classes at a community college. While transfer students are around 40% of the incoming degree-seeking undergraduate 12 month enrollment in the US (NCES, 2020), there has been limited focus on them as a group among chemistry education researchers (Wesemann, 2005) outside of a few initiatives such as the t-STEM program (Jewett *et al.* 2018). Transfer students more frequently have employment in addition to their coursework while also being more likely to commute to campus than their peers, which has been suggested as a potential reason for their lower rates of success in general chemistry relative to their peers (Stitzel and Raje, 2021). As outside employment and commuter status limit the potential for transfer students to socialize among their peers, considering transfer students specifically when investigating a social learning technique like PLTL can provide valuable insights into how the program is serving various populations of students including those who may not have access to all of the intended benefits that could extend outside the classroom.

Another way that students come into PLTL groups with different experiences can be attributed to the various demographic categories to which they belong. Particularly when considering features such as motivation, there have been reports of significant differences among groups. For example, it has been reported that female students typically have lower motivation than their peers in general chemistry (Liu *et al.*, 2017). Additionally, there are reports showing how motivation is correlated to early college academic success among Hispanic students (Kaufman *et al.*, 2008) while other reports demonstrate Hispanic students pass general chemistry at a rate lower than their peers (Mason and Mittag, 2001). When considering how students interact with each other in a PLTL setting, it is valuable to know how these differences are either increased, maintained, or decreased as a result of social influence in order to better plan activities to promote student learning and motivation.

We had an interest in understanding how PLTL at a particular university supports developing students' motivation in chemistry. Further, we wanted to understand more about how PLTL was serving the various demographic subgroups of the course to ensure that no group was being left behind. However, we were concerned about how the relational and non-independent nature of students in this setting might impact the validity of any claims made *via* traditional statistical methods due to the increased chance of false positive results (Scariano and Davenport, 1987). We also wanted a plan of

1004 | *Chem. Educ. Res. Pract.*, 2023, **24**, 1003–1024

This journal is © The Royal Society of Chemistry 2023

analysis that would let us understand more about how students are affecting each other as a result of their interaction. Therefore, in order to investigate motivation in a particular PLTL setting, we chose to use social influence models, a technique from SNA similar to multiple regression (Leenders, 2002). By using these social influence models, we seek to better understand student motivation in this context along with nuances of how students can affect each other. Additionally, we wish to use this project to demonstrate to the chemistry education research community a way of considering relational data that can help elucidate features of social learning environments that are not obtainable through traditional statistical methods.

Theoretical framework

Self-determination theory

Motivation is what helps to energize and direct students to reach outcomes such as academic achievement (Wigfield and Cambria, 2010). Self-determination theory (SDT) provides a framework for understanding student motivation in a way that is concerned with social environments that could enhance or prevent the development of motivation (Deci and Ryan, 1985). Within the macro theory of SDT are a number of mini-theories that take a more granular approach to understanding motivation. Among these mini-theories is Basic Needs Theory (BNT) which suggests that individuals need to fulfill basic needs of autonomy, competence, and relatedness to promote intrinsic motivation (Ryan and Deci, 2000) as illustrated in Fig. 1. Within BNT, autonomy refers to people's desire to independently make decisions for their own actions (Deci and Vansteenkiste, 2004). Being autonomous does not mean that one acts without regard to others, but that the decisions made are in accordance with the individual's interests and values. In the context of students, autonomy can refer to students' ability to make choices about

the courses they take or how they approach assignments within a course. Next, competence refers to an individual's ability to interact effectively with a particular environment (White, 1959). So in the context of students in a classroom, competence refers to their ability to successfully meet the learning objectives and perform well at evaluations. Finally, relatedness refers to the tendency for individuals to desire forming and maintaining interpersonal attachments that endure (Baumeister and Leary, 1995). Within the classroom, students have the opportunity to interact with peers and form friendships that can contribute to a sense of relatedness.

Within chemistry education research, SDT has been used as a framework for understanding motivation in a variety of settings. This research includes finding a positive relationship between student motivation and visuospatial skills in the context of learning group theory in a spectroscopy course taught by process oriented guided inquiry learning in Australia (Southam and Lewis 2013). Another setting where SDT was utilized was in Juriševič et al. (2012) when they found that students with higher motivation outperformed their peers in activities related to visible spectrometry.

SDT has also been used as a foundation when designing activities to support student learning and motivation (Ferreira et al., 2022; Wellhöfer and Lühken, 2022; Williams and Dries, 2022). Ferreira et al. (2022) designed inquiry-based laboratory activities for Brazilian high secondary students. The activities promoted autonomy (students approached activities as they saw fit), competence (students developed skills to propose procedures for experiments), and relatedness (these activities required teamwork). Through questionnaires, student interviews, and teacher's observations in a logbook, Ferreira et al. found evidence to support that the activities promoted intrinsic motivation. Wellhöfer and Lühken (2022) reported on a laboratory course designed around problem based learning. From student interviews, they found giving students the ability to autonomously propose experimental procedures was helpful in supporting student motivation. Finally, Williams and Dries (2022) reported on a guided-inquiry laboratory course for intermediate level chemistry majors (bioanalytical). From survey data, Williams and Dries found that many students attributed the ability to approach experiments autonomously or social factors as features that were helpful to their learning.

In addition to the previous examples of activities designed explicitly around SDT, the design of PLTL has many elements that would support student motivation according to the principle of fulfilling the basic needs laid out by SDT (Liu et al., 2018). For autonomy, students are not given a strict way to approach problems and peer leaders are instructed to support students in approaching the activities as the students see fit. For competence, the opportunity to work on the activities with the PLTL session is intended to enable students to gain more confidence in their ability to perform other chemistry problems. For relatedness, students' interactions with their peers provide opportunities for forming lasting relationships with peers that can extend even beyond general chemistry.
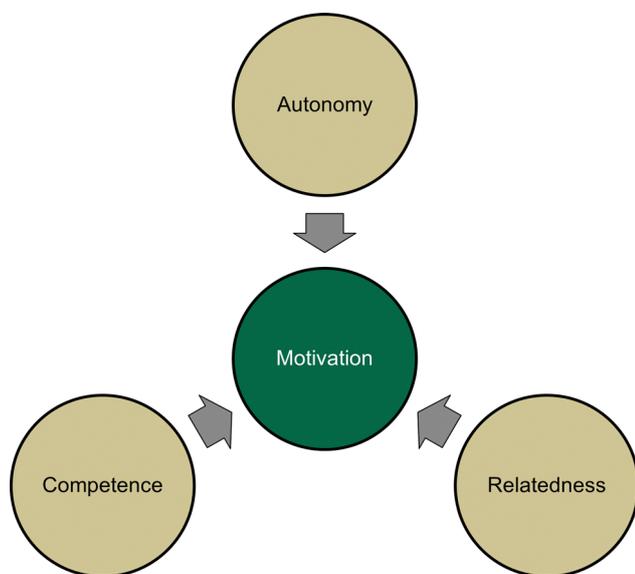

Fig. 1 Illustration of the three basic needs which need to be satisfied to promote motivation according to basic needs theory (BNT).

This journal is © The Royal Society of Chemistry 2023

Chem. Educ. Res. Pract., 2023, 24, 1003–1024 | 1005

Social comparison theory

In order to gain a better self-evaluation of one's own opinions or abilities, individuals will compare themselves to other individuals. A framework to understand this behavior, social comparison theory, was first proposed by Festinger (1954) which laid out nine main hypotheses that have been elaborated and expanded upon since the initial proposal, and this theory has found its way into many research studies of elementary and secondary schools (Dijkstra et al., 2008). One idea that comes from this work is the idea that people are more likely to adjust their opinions or seek to improve their abilities based on comparisons with people who are generally similar to themselves (Hypothesis 3, Festinger, 1954). Within PLTL, the use of the peer leader is valuable because it provides students a model who is similar to themselves, unlike how a faculty member might be perceived (Gosser and Roth, 1998). However, while abilities experience a generally unidirectional drive to be improved, opinions do not operate on such a clean directional path (Hypothesis 4). Generally, comparisons are categorized as upwards (comparison to someone perceived better off), downwards (comparison to someone perceived worse off), or lateral (comparison to someone perceived at a near equal level). PLTL should facilitate students making an upwards comparison to their peer leader and feeling a desire to improve their abilities while also matching the peer leader's positive opinions toward chemistry and learning. In addition to their peer leaders, students will be making comparisons to their peers as part of the PLTL learning process.

One way that this social comparison can manifest is when students could become more similar to each other through 'normative processes' (Dijkstra et al., 2008). In this case, students' opinions or abilities will become more similar to their peers as a result of interaction (Felson and Reed, 1986). These processes would result in students who are above average on some outcome going down toward the average, and students below average approaching the average. An additional type of comparison is some students might see a high performance among their peers as a source for a negative self-evaluation as they do not meet the same standard as their peers. Simultaneously, the students with high performance will have that high performance as a standard to support that they are in fact high performing. These two behaviors create a feedback loop or 'contrast effect' which causes students to diverge in their outcome measures such that the range of the outcome measure spreads as a result of the social influence. Generally, many findings in education research looking at social influence of self-concept show a contrast effect (Dijkstra et al., 2008).

Within chemistry education research, social comparison theory was directly used as a framework to analyze how general chemistry students engaged in a simulated peer review activity (Berg and Moon, 2022). In this study, Berg and Moon found that students were motivated by a desire for self-improvement where the students would take what was good in the reviewed responses to improve their own. This is in contrast to a self-enhancement motivation where students gain confidence in their existing response based on a downward comparison to someone with perceived lower ability.

Research questions

As the statistical social influence models we are using are not reported in the chemistry education literature to the best of our knowledge, we ran the analysis with more conventional methods to serve as a point of comparison. This starting point allows us to precisely discuss the unique outcomes afforded by the social influence models that would not be accessible via traditional analysis techniques while steadily building our understanding of the data and results. Additionally, as social influence can only occur as social interactions happen, and we are aware there is variation in how much students interact with specific peers, we also wanted to investigate if the amount of interaction between pairs of students might have an effect on the amount of influence experienced in our data. To guide our investigation, we laid out the following research questions:

1. To what degree do three student groups of interest (transfer, female, and Hispanic) differ from their peers within our sampled population on measures of motivation, autonomy, competence, and relatedness?

2. To what degree does interest in PLTL activities follow the pattern described by Basic Needs Theory (BNT) and illustrated in Fig. 1 for general chemistry students within our sampled population?

3. What patterns do we observe in the social influence of motivation toward the end of the semester, and what does social comparison theory suggest about these patterns?

4. To what degree does the amount of social interaction over a semester relate to the intensity of social influence?

Methods

The data collection for this study was approved by the Institutional Review Board at the University of South Florida and consent was obtained for all responses used in analyses. All the data analysis in this study was conducted on R version 4.1.1 (R Core Team, 2021).

Research setting

The students who participated in this study were enrolled in the first semester of a general chemistry sequence at a large southeastern university in the Fall semester of 2021. The nine sections of the course were coordinated across four instructors and followed a common syllabus. For assessment, the students in all sections took a common exam at the same time. To support student learning, the department has implemented a PLTL program in the general chemistry sequence. So in addition to attending two 75 minute lectures per week, students are assigned into a PLTL section that meets on Fridays. Attendance at these sessions is considered part of the overall course grade. Each PLTL section serves 18–23 students and is assigned a peer leader who facilitates two sections each. The time of a student's

1006 | Chem. Educ. Res. Pract., 2023, **24**, 1003–1024

This journal is © The Royal Society of Chemistry 2023

PLTL section on Friday is determined at registration; however, students are randomly assigned into particular sections by the peer leading coordinator. Within the PLTL sections, peer leaders are instructed to make groups of 3–4 students that are balanced between female and male students to the best of the peer leader's ability. Groups are otherwise not designed based on any other student demographic. For a particular session of PLTL, peer leaders are given discretion to move students or combine groups to help maintain the group size of 3–4, but otherwise groups are intended to remain consistent throughout the semester.

The particular semester for this study, Fall 2021, was unique in terms of how it was affected by COVID-19. At the beginning of the semester, all class activities were planned to be in-person. However, whenever a student tested positive for COVID-19, the section of the course and associated PLTL sections were moved online for a 2 week period. Shortly before the first exam, the decision was made that due to the need for direct interaction among students in PLTL, the Friday sessions would be online for the remainder of the semester while the lecture would remain in-person when possible. Online PLTL sessions were conducted in Microsoft Teams both before and after the decision to conduct all sessions online. For online sessions, students met in a Teams room assigned to their particular PLTL section, were split into breakout rooms for group work, and came together as a class to discuss answers after working in their particular group.

Instrument selection and adaptation

In order to answer our research questions, we needed to operationalize each of the components of BNT in order to measure them and understand their relationship with each other. Based on previous research (Southam and Lewis 2013; Liu, 2017) along with availability of instruments (McAuley *et al.*, 1989; Williams and Deci, 1996), we proceeded in the following way: for autonomy, there were a few challenges due to the lack of potential for students to approach general chemistry in a fully autonomous way. Most students take general chemistry as part of fulfilling requirements for their degree, and relatively few students are specifically chemistry majors (Lloyd and Spencer, 1994). Additionally, within general chemistry, students are generally not given the option to create or choose their own assessments outside some innovative practices such as "Cafeteria-style" grading (Goodwin and Gilbert, 2001) which are not being implemented in our context. Considering these limitations, the particular way we approached autonomy in this study is through autonomy support. The idea of autonomy support in our context will relate to how students perceive their peer leaders as supporting their ability to approach the activities of a particular session in any way they see fit and not simply following the 'correct' path to an answer. In order to characterize autonomy support, we used the Learning Climate Questionnaire (LCQ; Williams and Deci, 1996) as a 15-item 7-point scale. The items of the LCQ were presented in the same order as previous administrations. This scale has previously been used to characterize students in

organic chemistry classrooms to show that autonomy support predicted course performance (Black and Deci, 2000).

For the other three components related to BNT (motivation, competence, and relatedness), we chose to use measures from the Intrinsic Motivation Inventory (IMI; McAuley *et al.*, 1989). The IMI was developed to include a number of potential scales that researchers can mix and match to their particular research questions to approach many aspects of SDT. For our research, we chose to use scales for *interest*, *perceived competence*, and *relatedness* to characterize motivation, competence, and relatedness respectively. These three scales of the IMI have been used in previous reports using SDT to analyze motivation in chemistry classrooms (Southam and Lewis 2013; Liu, 2017; Ferreira *et al.*, 2022). The scale of *interest* is considered to be the self-report measure on intrinsic motivation (McAuley *et al.*, 1989) and allows us to efficiently gain an understanding on motivation while simultaneously exploring other constructs. For competence, we used perceived competence as the self-report impression that students are able to successfully answer the prompts in the activities within their PLTL session. For relatedness, the relatedness scale of the IMI gives us a measure about each student's desire to form and maintain interactions with the other members of their PLTL group.

A few adjustments were made to adapt the instruments to our setting. First, the items of the LCQ were presented with the original text of "my instructor" being replaced with "my peer leader". This change in wording was made to focus the items on the peer leaders in terms of how they are supporting student autonomy. After completing the LCQ, students were presented items from the IMI. While setting up the instruments, we randomized the order of items in the IMI such that items from all scales were shuffled among each other, but each student would see the same items in the same order. For the IMI, the phrasing of "this activity" was replaced with "Peer Leading Activities" along with minor adjustments to the subject-verb agreement and "this person" was replaced with "my peer leading group" to match the context. These choices in wording were decided based on discussion with the peer leading coordinator to best match how the activities within the PLTL sessions are usually described to students by their peer leaders. The items for these instruments are presented in Appendix A.

Data collection

The instrument for characterizing student motivation as related to SDT and specifically BNT was administered *via* Qualtrics at two points during the semester. These two time points were chosen in order to see how these constructs develop over the semester. The first round of data collection took place during the 4th week of the semester and closed the day before the first exam. This week was chosen to be early in the semester but still give students the opportunity to attend a few PLTL sessions so the items would be meaningful to them. The second round of data collection occurred in the 13th week of the semester which was after the third exam. During the first round of data collection, students had attended up to two PLTL

sessions with some being in-person and some being online. All PLTL sessions after the first exam occurred exclusively online. The instrument was announced on the course management system, Canvas, where students were provided a link to take the instrument. To improve participation, reminder messages were also sent. As an incentive for completion, students were awarded 5 bonus points for completing the instrument at each time point which was added into their course total of 1000 points. Students were given the opportunity to consent to having their responses used for research before each administration. Students were given the opportunity to earn the points for completing the instrument regardless of whether or not they consented to be included in the research data. Student demographic information was obtained from university records. This data included admission type which was used to classify students who were non-transfer (first time in college), transfer (either from the state community college system or outside), or other (which includes non-degree seeking students).

Data preparation began by removing all responses from students who did not consent to having their responses analyzed for research purposes. Responses were then checked against course enrollment to match responses to other student data. Each case where a student gave an identical response to every item on either the LCQ or IMI was removed from the analysis in order to improve data quality. This choice was made because, as the instrument included a mixture of positively and negatively phrased items, an identical response to every item suggests the respondent was not thoughtfully reading and responding to the particular items. This straight lining behavior has also been associated with participants speeding through responses in order to finish the instrument without thoughtfully engaging with the items as they go (Zhang and Conrad, 2014). The last data preparation step was to reverse code all items which were negatively phrased to aid in interpreting results.

There were a total of 1988 students enrolled in the course. After data processing and cleaning, we had a total of 1179 (59% of total students) responses from the first administration and 1244 from the second administration (63%). Analyzing response rates by demographics (Appendix B, Table 8) did not reveal any concerning trends in response rate across demographic groups, which might indicate evidence of a response bias and therefore merit a change in the analysis plan.

As part of running a social influence model, we needed to define the social network through which students were connected. As part of their responsibilities, peer leaders recorded which students attended each session as well which students worked together during the sessions. From this attendance data, we constructed social networks based on the group composition for each particular session by considering members within a particular PLTL group as having undirected ties among each other. These particular social networks were then further refined based on the particular research question being addressed.

Internal structure validity

Before we approached our research questions, we analyzed our instrument for evidence of validity (Arjoon *et al.*, 2013; AERA,

APA & NCME, 2014; Lewis, 2022). Specifically, we checked for aspects of internal structure validity using exploratory factor analysis, confirmatory factor analysis, and measurement invariance testing. For these analyses, we split our data into two halves using a random sort technique in order to have separate data sets of equal size for each analysis. The highlights of these analyses will be presented here while more detail is provided in Appendices 3–5. The highlights of these analyses are as follows: the exploratory factor analysis (Appendix C) from our initial 4 scales (Autonomy Support, Interest, Perceived Competence, and Relatedness) suggested that we ought to remove negatively phrased items, and that the remaining instrument was either 3 or 4-factors. While the decision to remove reverse phrased items was primarily based on the results of the EFA, many of these items were found to be problematic in previous administrations. For example, lcq.13.r has been removed due to its low loading, and several reverse phrased IMI items (imi.02.rel.r, imi.06.rel.r, imi.11.int.r, and imi.18.rel.r) have been removed based on examination of modification indices (Liu, 2017). Therefore, there is some precedent for removing the same items that we did. The 3-factor solution generally divided the items into factors of Autonomy Support, Perceived Competence, and a factor that included both Interest and Relatedness items. The potential 4-factor solution was similar but resolved the items characterizing Interest and Relatedness into their respective factors. After these results from exploratory analysis, subsequent confirmatory factor analysis (Appendix D) on the reserved portion of the data supported that the 4-factor solution (χ^2 (344, $N = 623$) = 1692, SRMR = 0.064, RMSEA = 0.079, and CFI = 0.896) was a better fit than the 3-factor solution (χ^2 (347, $N = 623$) = 1919, SRMR = 0.067, RMSEA = 0.085, and CFI = 0.878) with the fit statistics presented being from the second administration (see Appendix D for full discussion of fit statistics including fit statistics from first administration). Considering the match with our intended theoretical constructs from the instrument design and improved fit statistics of the 4-factor model, we proceeded with the 4-factor solution. In the interest of exploring alternate measurement models that utilized more of our collected data, we made several attempts to model all items in the instrument instead of only the positively phrased items. Initially, we ran a CFA that included all items based on their intended theoretical scales. This CFA was followed by another model that incorporated a negative 'methods' factor, which is a technique that has been shown to improve model fit (Naibert and Barbera, 2022). Finally, we tested a bifactor model, as that also might have provided a way to better understand our data. Since the fit of none of these models represented a meaningful improvement, we elected to retain the four-factor model without reverse phrased items, as this model is the most straightforward to interpret.

Measurement invariance testing (Appendix E) was then conducted in order to check if we had evidence that the internal structure was consistent between the two administrations and across some key demographic groups. This step is important because if measurement invariance does not hold, then the constructs cannot be meaningfully compared as they potentially

1008 | *Chem. Educ. Res. Pract.*, 2023, **24**, 1003–1024

This journal is © The Royal Society of Chemistry 2023

have different internal structures across time or demographics (Chen, 2007; Rocabado *et al.*, 2019; Rocabado *et al.*, 2020). When comparing the first and second administration, we found evidence of strict invariance supporting that we can directly compare scores produced by the two administrations without need for further adjustments. Additionally, we observed strict invariance between female and male, white and Hispanic, and non-transfer and transfer students suggesting that we can use the factor scores from the instrument to directly compare students in these pairs of groups.

Social influence modeling

In order to analyze our data, we wanted to use a method that is able to consider interactions among individuals. By incorporating interactions among individuals in our analysis, we can address concerns related to Type I error (Scariano and Davenport, 1987; Stevens, 2007) while also potentially elucidating mechanisms for how students are influencing each other. Specifically, we analyzed our data using social influence models (Lane *et al.*, 2019; Leenders, 2002). Within the social influence modeling strategies, there are two ways to describe social influence: network effect (auto-regression) and network disturbance (moving average). The mathematical formula for a network effect is $y = \rho W y + X\beta + \varepsilon$. While a full description of how each term is constructed is beyond the scope of this paper, we will present the interpretation of each term and recommend that interested readers refer to Leenders (2002) for expanded details. The terms in the equation are as follows: variable y is a vector of the outcome measure. For $\rho W y$, the network effect term ρ is a weighting constant, W represents the social network through which influence occurs using a normalized 2-dimensional adjacency matrix, and y is again the vector of the outcome measure for all students. The $X\beta + \varepsilon$ portion of the equation is directly analogous to multiple regression, and in fact, the network effects equation becomes a multiple regression when there is no social influence ($\rho = 0$) or when individuals are considered entirely independent (every element of the matrix W is 0). For X, β, and ε, X is the matrix of measured predictors and an intercept term, β is a vector of weighting coefficients, and ε is a residual term. The practical implication of a significant network effect, ρ, depends on the sign and magnitude as illustrated in Fig. 2 where the three circles represent three individuals, the shading of each circle represents the value of the outcome measure, and lines between circles represent interaction between individuals. As mentioned before, if $\rho = 0$, then the network effects equation becomes identical to a multiple regression, and there is no change in the outcome measure as a result of student interaction. If $\rho < 0$, then that is evidence that individuals are likely to become more polarized on the outcome measure as a result of interaction. This tendency is illustrated in Fig. 2 by increasing the differences in the shading of the circles suggesting these individuals have become less similar as a result of inter-action. For situations where $\rho > 0$, that would suggest that individuals are more likely to become more like each other on the outcome measure as a result of prolonged interaction. This idea is illustrated in Fig. 2 by having the circles match each other in shade. It should be noted that while Fig. 2 might imply that the
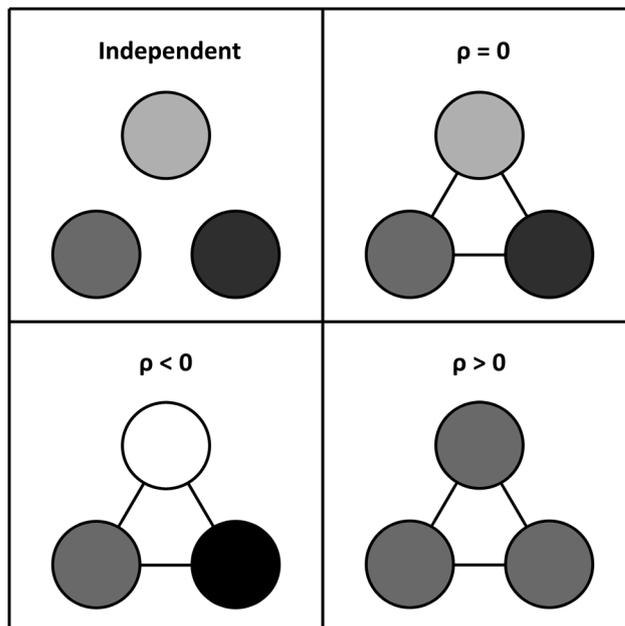

Fig. 2 Graphical representation of individual outcome measures at different levels of network effects. Each circle represents an individual and a line between circles indicates social interaction. Each quadrant represents a group of individuals that have identical baseline levels of an outcome measure before controlling for network effects. The value of an attribute is suggested by the shading in each circle.

average value for an outcome is maintained during a social influence process, this is not necessarily what will be observed in practice, and this figure is merely meant to suggest the relative behavior within groups.

Another form of social influence, network disturbance, is described by the pair of equations $y = X\beta + \varepsilon$ and $\varepsilon = \rho W\varepsilon + \upsilon$. What this mathematical model suggests is that the residual terms of the regression are influenced by the network. For the classroom setting, this kind of finding would imply that students who interact rise or fall together controlling for other factors in the regression. The first equation, $y = X\beta + \varepsilon$, is identical to multiple regression and to how the terms were defined in network effects. The additional contribution for network disturbances will be that the residual term is going to be influenced by the social network. So for $\rho W\varepsilon + \upsilon$, ρ is the weighting constant for the network disturbance, W represents the social network as it did in network effects, ε represents the residual terms from the regression, and υ is the remaining residual after accounting for the amount explained by the network influence.

Model specification

As we planned to use a novel method of data analysis, we decided to include two conventional analyses to serve as a point of comparison for the social influence model. When planning our analyses, we chose to compare three models which we named Model 1, Model 2, and Model 3 with the similarities and differences among the models summarized in Table 1. All three models follow the same general pattern of using data from the

This journal is © The Royal Society of Chemistry 2023

Chem. Educ. Res. Pract., 2023, **24**, 1003–1024 | **1009**

second data collection to predict interest as a function of autonomy support, perceived competence, and relatedness. Additionally, the value for interest at the first data collection is included in each model as a covariate. One distinction among the models will be how the factor scores are computed. For Model 1, we computed 'coarse' factor scores by taking the unweighted mean of the responses for items in a factor. These coarse factor scores have the advantage of being simple to calculate and simple to interpret as the produced values are directly analogous to the response options available when answering items. However, these coarse scores are unable to parse out correlations among factors and contributions from measurement error (Brown, 2015). Therefore, we used 'refined' factor scores for Model 2 and Model 3 where these values are computed using the measurement model described in the section on internal structure validity. Refined factor scores were calculated using the lavPredict fucntion in the lavaan package in R (Rosseel, 2012) using the method "regression". Beyond factor score computation method, Model 3 is distinct as it considers the impact of social influence by either network effects (ρ_1) and network disturbances (ρ_2) processes in addition to the other terms. All models were computed in R using the *lnam* function which utilized the *sna* (version 2.6) library (Butts, 2020). For Model 1 and Model 2, the networks were left unspecified in order to get the multiple regression results.

Before we ran our multiple regression and social influence models, we needed to determine which students would be included in the analysis and how we would establish which students were connected. At first, we used the group composition from the session before the second data collection was opened as our social network. As peer leaders were instructed to keep students together to the best degree possible and this particular session occurred late in the semester, we believed this particular session would be representative of students who had significant interactions with each other and had the most opportunities for social influence. Therefore, this was the network we used when running the Model 3 that is directly compared to Model 1 and Model 2.

However, as we also wanted to explore how social influence varied by the amount of interaction among students (research question 4), we needed to establish a different set of networks to address this question. For this task, we started by creating a weighted network object where the weight of a tie between two individuals was the count of the number of times the peer leader recorded the pair being in the same PLTL group. So even if groups were fluid based on the attendance for a particular session, this weighted network would be able to identify pairs of students who worked together consistently even if other students in the group varied week to week. From this weighted network object, we created a series of unweighted networks by restricting the network to only include ties that reached a certain tie weight threshold. As an example, the students in the network with the weight threshold of 3 include pairs of students who interacted with each other at least three times though certain pairs could have interacted more than three times. So low weight thresholds result in networks predominantly composed of interactions that were not maintained throughout the semester while high weight thresholds result in networks predominantly composed of interactions that persisted throughout the semester.

From these networks, we needed to further restrict the particular students who would be included in our statistical models. For starters, as we wanted to use data from both timepoints in the analysis, we eliminated students who did not take and consent to both administrations of our instrument. Additionally, as we wanted to consider the impact of social influence on the students, we did not want to include students for whom we were lacking data from their group members as this lack of data could complicate how we interpret social influence. To address this concern, we removed students for whom we did not have complete data from at least 2 group members.

We started with 1179 students who completed the initial instrument and 1244 who completed the final instrument. These numbers take into account students completing the instrument, consenting to data analysis, and not straight lining their responses. In total, we had a total of 870 students who completed both the first and second data collections. Then, after filtering students for whom we did not have complete data from at least 2 peers, we were left with 270 students to analyze. Tables comparing the 270 included students to the rest of the sample are presented in Appendix F (Tables 20–22). For the separate set of network objects defined by the number of interactions among students, we started with 735 students who completed both data collections and interacted at least 1 time with at least 2 others who also completed both data

Table 1 Summary of models to be compared. Coarse factor scores are the unweighted mean of responses to items within a factor while refined scores are computed using the measurement model

	Model 1	Model 2	Model 3
Factor score computation method	Coarse	Refined	Refined
Social influence terms	N/A	N/A	Network effects (ρ_1) and network disturbance (ρ_2)
Outcome measure		Interest (time 2)	
Predictors		Interest (time 1) Autonomy support (time 2) Perceived competence (time 2) Relatedness (time 2)	

Chemistry Education Research and Practice

collections. As the threshold increases for interaction increases, the number of students included decreases. For example, from the data we can model, 370 students interacted with peers at least 3 times, 105 interacted at least 7 times, and 24 interacted 11 times.

Results

Score comparison

Before investigating the results of our models, we began by looking at our data more descriptively. This step helps us better understand the student population and better interpret any implications from the subsequent statistical analyses. Our first task was to compare the coarse factor scores across the two data collections and across available demographic categories with these scores being summarized in Tables 2–5. These scores are calculated on a 1 to 7 scale with 1 representing the strongest negative response, 4 being a neutral response, and 7 being the strongest positive response. As was mentioned in the section on internal structure validity, we observed strict invariance between the two data collections (longitudinal invariance), transfer and non-transfer students, female and male students, and White and Hispanic students. This observation of strict

Table 2 Summary of unweighted "Interest" factor item averages across demographics. Items are on a 1–7 scale with 4 being a neutral response

Students	Administration	n	Mean	sd	Cohen's d
All students	Time 1	1179	4.18	1.40	
All students	Time 2	1244	3.74	1.52	0.30
Non-transfer	Time 1	1037	4.14	1.38	
Transfer	Time 1	107	4.36	1.52	0.16
Non-transfer	Time 2	1096	3.72	1.51	
Transfer	Time 2	113	3.66	1.58	−0.04
Female	Time 1	759	4.17	1.40	
Male	Time 1	420	4.19	1.41	0.02
Female	Time 2	762	3.66	1.53	
Male	Time 2	482	3.85	1.50	0.12
White	Time 1	475	3.99	1.39	
Hispanic	Time 1	261	4.22	1.36	0.17
White	Time 2	507	3.61	1.50	
Hispanic	Time 2	270	3.77	1.44	0.11

Table 3 Summary of unweighted "Autonomy Support" factor item averages across demographics. Items are on a 1–7 scale with 4 being a neutral response

Students	Administration	n	Mean	sd	Cohen's d
All students	Time 1	1179	5.56	0.93	
All students	Time 2	1244	5.64	0.99	−0.08
Non-transfer	Time 1	1037	5.56	0.92	
Transfer	Time 1	107	5.57	1.05	0.01
Non-transfer	Time 2	1096	5.63	0.99	
Transfer	Time 2	113	5.72	0.99	0.09
Female	Time 1	759	5.56	0.96	
Male	Time 1	420	5.56	0.87	0.00
Female	Time 2	762	5.66	1.01	
Male	Time 2	482	5.61	0.96	−0.05
White	Time 1	475	5.56	0.94	
Hispanic	Time 1	261	5.53	0.92	−0.03
White	Time 2	507	5.65	1.01	
Hispanic	Time 2	270	5.59	0.97	−0.03

Table 4 Summary of unweighted "Perceived Competence" factor item averages across demographics. Items are on a 1–7 scale with 4 being a neutral response

Students	Administration	n	Mean	sd	Cohen's d
All students	Time 1	1179	4.80	1.30	
All students	Time 2	1244	4.74	1.36	0.05
Non-transfer	Time 1	1037	4.83	1.29	
Transfer	Time 1	107	4.58	1.37	−0.20
Non-transfer	Time 2	1096	4.78	1.35	
Transfer	Time 2	113	4.37	1.32	−0.31
Female	Time 1	759	4.67	1.35	
Male	Time 1	420	5.04	1.18	0.28
Female	Time 2	762	4.62	1.42	
Male	Time 2	482	4.93	1.24	0.23
White	Time 1	475	4.73	1.28	
Hispanic	Time 1	261	4.77	1.31	0.03
White	Time 2	507	4.77	1.38	
Hispanic	Time 2	270	4.61	1.25	−0.12

invariance gives support for making direct comparison across time points and between various demographic categories (Chen, 2007; Rocabado *et al.*, 2019; Rocabado *et al.*, 2020). In order to better contextualize the differences observed in Tables 2–5, values for Cohen's d were calculated to provide an understanding of the effect size of the difference. Traditionally for Cohen's d, values of 0.2–0.5 are considered 'small' effects, 0.5–0.8 are considered 'medium', and >0.8 are considered 'large' (Cohen, 1988).

For all students, we see there is a drop in interest (Table 2), perceived competence (Table 4), and relatedness (Table 5) as the semester progresses. By Cohen's d, the drop in interest is small, the drop in perceived competence is trivial (<0.2), and the drop in relatedness is large. Of note is that while the drop in interest is small, it does cross the threshold of going from a positive score (4.18) to a negative score (3.74) considering the neutral score of 4. This suggests that while students started with a generally positive interest in activities within their PLTL sessions, this was not maintained throughout the semester. Additionally, the relatedness score experienced a similar but much larger drop across the two administrations from 5.44 to 3.48. This finding suggests that across the two time points,

Table 5 Summary of unweighted "Relatedness" factor item averages across demographics. Items are on a 1–7 scale with 4 being a neutral response

Students	Administration	n	Mean	sd	Cohen's d
All students	Time 1	1179	5.44	1.06	
All students	Time 2	1244	3.48	1.38	1.59
Non-transfer	Time 1	1037	5.44	1.04	
Transfer	Time 1	107	5.44	1.18	0.00
Non-transfer	Time 2	1096	3.46	1.37	
Transfer	Time 2	113	3.56	1.46	0.08
Female	Time 1	759	5.43	1.09	
Male	Time 1	420	5.46	0.99	0.02
Female	Time 2	762	3.37	1.41	
Male	Time 2	482	3.64	1.33	0.19
White	Time 1	475	5.42	1.07	
Hispanic	Time 1	261	5.40	1.04	−0.02
White	Time 2	507	3.38	1.32	
Hispanic	Time 2	270	3.40	1.33	0.01

This journal is © The Royal Society of Chemistry 2023

Chem. Educ. Res. Pract., 2023, **24**, 1003–1024 | 1011

students went from a generally positive desire to form and maintain relationships with their peers to a generally negative desire. In contrast, while perceived competence dropped trivially between the administrations, it maintained a consistently positive value (4.80 to 4.74) suggesting students consistently felt positive about their ability to succeed in doing the activities within their PLTL sessions. While many of the factor scores dropped as the semester progressed, we observed a slight rise in autonomy support but with a trivial effect size (Table 3). Both coarse averages of autonomy support at the two time points (5.56 and 5.64) were positive relative to the neutral value of 4 suggesting students consistently felt their peer leaders supported them in approaching the activities within their PLTL sessions as the students desired.

When comparing various demographic categories, we see that the differences often have quite small effect sizes. In fact, no comparison for interest, autonomy support, and relatedness for either administration reaches the traditional threshold for a small effect size of 0.2 when looking at Cohen's d. In contrast, we see some small effect size differences when comparing the perceived competence of non-transfer/transfer students and with female/male students. For the comparison between non-transfer and transfer students, transfer students start with a lower perceived competence and this gap increases as the semester progresses. Then for females and males, males start with a higher perceived competence relative to their peers though this gap does shrink slightly between the two time points. In the cases where there is a small effect size difference in perceived competence, the difference does not represent a change from positive to negative perceived competence.

Statistical models

In addition to looking at the descriptive statistics of the factor scores from our data, we wanted to understand more about the composition of groups that would be included in our models. Our social network data for students who met our criteria for inclusion in the multiple regression and social influence analyses included 270 students in 82 groups. The group sizes have various numbers of students within a small range including 3 ($n = 61$), 4 ($n = 19$), 5 ($n = 1$), and 6 ($n = 1$) students. Of the 82 groups, not a single group contained more than 1 of a total of 19 transfer students. Also, we see that 20 groups were all female while none were all male with 17 of those 20 being within groups of 3. Having some all-female groups and no all-male groups is not surprising as 67% of the sample is female. For White and Hispanic students, there were 5 groups which were all White and 1 group which was all Hispanic.

When analyzing student interest using multiple regression (Model 1 and Model 2) and a social influence model (Model 3), we first wanted to see evidence that BNT is a predictive framework followed by seeing how each model added to our understanding of the data. The summary of the results from all of these models is shown in Table 6. From all three models, we see predictive power in all/most of our measured factors in predicting student interest. Specifically, initial interest, perceived competence, and relatedness are significant and positive predictors of interest. One

Table 6 Multiple regression summary table

Model		B	Std error	Z value	Sig	Fit
1	(Intercept)	−1.41	0.32	−4.425	<0.01[a]	$R^2 =$
	Interest ($t1$)	0.26	0.04	6.114	<0.01[a]	0.6807
	Autonomy support ($t2$)	0.20	0.07	3.071	<0.01[a]	
	Perceived competence ($t2$)	0.25	0.05	5.486	<0.01[a]	
	Relatedness ($t2$)	0.50	0.05	10.486	<0.01[a]	
2	(Intercept)	−0.10	0.05	−2.35	0.02[a]	$R^2 =$
	Interest ($t1$)	0.16	0.04	4.45	<0.01[a]	0.8017
	Autonomy support ($t2$)	0.07	0.06	1.22	0.22	
	Perceived competence ($t2$)	0.27	0.04	6.52	<0.01[a]	
	Relatedness ($t2$)	0.80	0.05	16.13	<0.01[a]	
3	(Intercept)	0.03	0.04	0.768	0.44	$R^2 =$
	Interest ($t1$)	0.15	0.03	4.430	<0.01[a]	0.8130
	Autonomy support ($t2$)	0.07	0.05	1.371	0.17	
	Perceived competence ($t2$)	0.26	0.04	6.534	<0.01[a]	
	Relatedness ($t2$)	0.82	0.05	16.834	<0.01[a]	
	Network effects (ρ_1)	−0.05	0.02	−2.163	0.03[a]	
	Network disturbances (ρ_2)	0.02	0.03	0.451	0.65	

[a] $p < 0.05$.

distinction between Model 1 and the others is that autonomy support falls below the conventional levels of statistical significance when using refined factor scores from the measurement model instead of the coarse score used in Model 1.

When looking at Model 3, the social influence model, we see that network effects is a significant negative predictor while network disturbances is not significant. The fact that network effects is significant and negative gives support that students are having an effect on each other, and this effect generally results in the range of student interest within a group expanding as a result of interaction.

Finally, we wanted to understand how the social influence was affected by the amount of interaction between students. To do this, we ran all the terms of Model 3 on the set of network objects we produced by varying the threshold of the number of interactions required to be included in the analysis. Fig. 3 represents how the network effect term varied based on the number of interactions
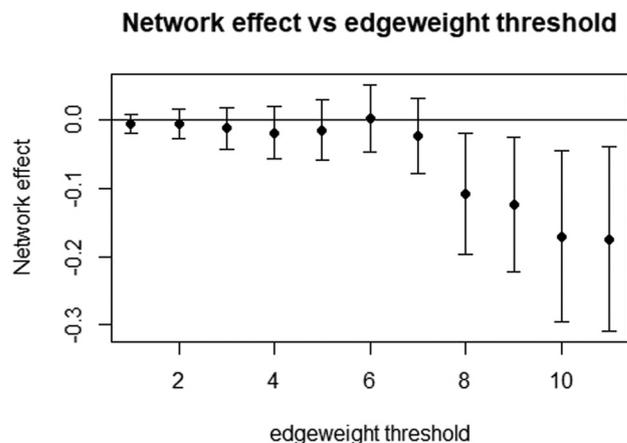

Fig. 3 Graph of how the network effect term from Model 3 can vary based on the amount of interaction between students. Bars from the observed value represent 95% confidence intervals.

This journal is © The Royal Society of Chemistry 2023

between students. From this data, we see that for low threshold values, we do not see a significant network effect term in the model (illustrated in the figure by the 95% confidence interval overlapping with zero). However, this behavior changes after the threshold of 8 interactions has been met. From there, we see the significant and negative network effect term like what was observed in the previous analysis that focused on the network toward the end of the semester. While the 95% confidence interval expands, partially as a result of fewer students being included in the analysis, the magnitude of the effect is large enough to be found significant. This finding suggests that while social influence is occurring between students, there is some threshold of interaction that needs to be reached before this influence is statistically observable.

Discussion

Demographic comparisons (research question 1)

For the first research question, we went through some descriptive analysis in order to better characterize groups of interest (transfer, female, and Hispanic) in our sample. Before looking at the particular groups of interest, we first looked at the data as a whole to better understand how the groups compare to the class. For all students between the first and second data collection, we saw a small drop in interest, a large drop in relatedness, and trivial changes in autonomy support and relatedness. The drops in interest and relatedness are particularly noteworthy as they went from above neutral (>4) to below neutral (<4) during the semester. While the drop in interest may be undesirable, it is not particularly surprising based on previous investigations of attitude for chemistry. This drop in attitude toward chemistry has been seen in previous studies of PLTL classrooms (Chan and Bauer, 2015; Liu *et al.*, 2017). It should also be noted that while we do observe a drop in interest, this should not necessarily be interpreted as a negative result of the PLTL program. First, our measure of interest specifically referred to the activities within the PLTL session which limits our ability to speak to a student's interest in chemistry as a general concept. Additionally, we only collected this data from students in PLTL so we do not have a point of comparison in this context to speak to how PLTL compares to traditional lectures. Many studies that have directly compared PLTL to traditional lecture have found a positive correlation between participating in PLTL and motivation relative to traditional lecture (Liu *et al.*, 2017; Frey *et al.*, 2018). Additionally, we do not have data that would allow us to make a direct comparison between PLTL delivered traditionally and the online delivery of PLTL.

Compared to the drop in interest, the drop in relatedness was quite stark with a large effect size. When analyzing this result, it is important to remember the context in which these data were collected. Specifically, this was a semester of PLTL that was intended to be in person that transitioned into an entirely virtual experience. Many CER studies reported findings such as the transition to online learning resulting in the loss of peer communication networks (Jeffery and Bauer, 2020).

Another reported finding is the observation that the online learning environment generally weakened student engagement due in part to working in settings that were not conducive to learning (Wu and Teets, 2021). Based on these previous findings, we believe the nature of the online interaction during this semester could play a strong role in the large drop of students' feeling of building and maintaining relationships with their peers in their PLTL groups. Additionally, the online delivery also limits students' ability to interact directly with their peer leader. As part of the design of PLTL is using peers due to their ability to more directly relate with their students than professors would be able to do (Gosser and Roth, 1998), this lack of interaction could have also led to some overall drop in students' feelings of relatedness.

After understanding the baseline for all students, we then looked at our groups of interest. There were three comparisons that we chose to explore based on our concerns with potential inequities and the availability of enough data to make a meaningful comparison: (1) transfer students (Wesemann, 2005; Stitzel and Raje, 2021), (2) female students (Liu *et al.*, 2017), and (3) Hispanic students (Mason and Mittag, 2001). As for the comparisons between time points, we focused the comparisons for these groups based on effect size instead of simple significance testing as simply commenting on significance can either lead us to focus on trivial differences that simply meet a threshold of statistical significance or ignore large differences that simply did not meet a threshold of statistical significance. We saw no evidence of difference among the demographic comparisons at either time point for interest, autonomy support, or relatedness that met the traditional threshold of a small effect size. However, we did see a small effect size difference within perceived competence at both timepoints for transfer and non-transfer students as well as female and male students. In particular, the gap for perceived competence between transfer students and non-transfer students expanded during the semester though remained a small effect size. Transfer students were generally isolated in their peer-leading groups so social comparisons would be more likely to occur through comparisons to non-transfer student peers. In contrast, the gap between female and male students shrank while also staying in the small effect size category. Female and male students were typically combined in most PLTL groups, as the peer leaders were explicitly instructed to avoid homogenous groups on this demographic which gave more opportunities throughout the semester for these groups to engage in social comparisons with each other.

Analysis of BNT in PLTL (research question 2)

The second research question we presented asked to what degree did BNT help describe interest in PLTL for our setting. According to BNT, fulfilling the basic needs of autonomy, competence, and relatedness can help develop motivation (Ryan and Deci, 2000). To approach this research question, we used multiple regression and social influence models to see the relationship between the elements of BNT with the outcome of interest. The results from Model 1, Model 2, and Model 3

This journal is © The Royal Society of Chemistry 2023

Chem. Educ. Res. Pract., 2023, **24**, 1003–1024 | **1013**

(shown in Table 6) are generally consistent with the expected patterns from BNT. In most cases, we see that there is a positive and significant relationship between interest and the factors of autonomy support, perceived competence, and relatedness which is what would be expected in a setting where BNT is acting as a predictive model. Therefore, we have data to support that teaching strategies which support these factors can have a positive impact on students' motivation. This kind of finding is well supported by the literature of designing interventions in chemistry classrooms around SDT (Ferreira *et al.*, 2022; Wellhöfer and Lühken, 2022; Williams and Dries, 2022) where promoting autonomy, competence, and relatedness were shown to have a positive correlation with student motivation.

However, the evidence in our models for autonomy support predicting interest is mixed. This observation comes from the fact that while autonomy support is a significant positive predictor for interest in Model 1, it is not significant in Model 2 and Model 3. The difference between Model 1 and the others is the method used in calculating factor scores. For Model 1, factor scores were 'coarse' meaning they were computed using the unweighted average of each item within the factor. This computational method is simple and is usable even when there are not enough students to run a reliable factor model. In contrast, Model 2 and Model 3 were 'refined' factor scores computed from the measurement model which can consider features such as items contributing unequally to the factor and measurement error. Therefore, one possible explanation of why this difference in significance is observed can be that the significance of autonomy support in Model 1 is an artifact of the coarse factor score treating each item in the scale as an equal contributor to the factor score. Regardless of the variation in significance for autonomy support between models, the way in which this course is structured relative to others might also affect the particular value of autonomy in this setting. A previous study has directly shown a relationship between autonomy support and success in organic chemistry (Black and Deci, 2000) and some contexts such as laboratories designed around problem based learning have suggested that giving students many options in how to approach a task is helpful to motivation (Wellhöfer and Lühken, 2022). In contrast, it might be the case in this particular setting that while students may strongly believe that their peer leaders will let them take any path they want to find an answer, they may still believe there is a 'correct' way to answer a particular question and that method should be determined. So while autonomy might be supported in this setting, students perceive the goal as finding that 'correct' method and do not necessarily see themselves as acting autonomously.

Social influence in interest (research question 3)

However, because our data is relational, we did not stop at multiple regression models due to the concern of a lack of independence (Theobald, 2018) and miscalculated standard errors (Scariano and Davenport, 1987; Stevens, 2007). Additionally, we are aware that students will compare themselves to their peers and make adjustments from that comparison as is described by social comparison theory (Festinger, 1954; Dijkstra *et al.*, 2008).

Therefore, we also asked Research Question 3 which dealt with asking about the types of social influence observed and how they related to social comparison theory. For this question, we will specifically look at our results from Model 3 where we can consider students as functioning within groups and the mechanisms by which they might be influencing each other (Lane *et al.*, 2019; Leenders, 2002). When comparing Model 2 and Model 3, we see that there is no difference in interpretation to the degree that you can interpret Model 2. So interest is still positively related to perceived competence, relatedness, and interest from an earlier point in the semester in both models. While we were concerned about how the relational nature of the data might affect these interpretations, Model 3 does not support that any of those conclusions change when considering the students as connected. What Model 3 was able to show was evidence of social influence in the data. This is a finding that is not even in principle possible with standard multiple regression. As social influence was observed, there is support that our concern that the data lacked independence was reasonable as we can see a statistically significant social influence occurring. While a lack of independence can have a strong impact on significance (Scariano and Davenport, 1987; Stevens, 2007), we did not observe any meaningful differences in significance between Model 2 and Model 3 among the terms that are shared between them.

Specifically, Model 3 gave support that our data exhibits a significant and negative network effect process. As the network effect term generally describes a process where an individual will be influenced by the group average (Leenders, 2002), a negative value here generally suggests a movement away from the average. While a full explanation of the model is more nuanced in how that happens due to the autoregressive nature of the model, the general observation that individuals become more polarized under a negative network effect holds. Based on social comparison theory, individuals look to others to better understand themselves (Festinger, 1954). In the case of some affective outcomes in educational settings, it is common to see that students are influenced by a contrast effect (Felson and Reed, 1986; Dijkstra *et al.*, 2008). Based on the observed negative network effect, our data seems consistent with an observation of a contrast effect which is what we believe is likely in play here.

In order to better illustrate the concept of network effects within our data, we found a group in the data to serve as an exemplar. The students in this group are assigned pseudonyms of Alice, Beth, and Charlie, and their interest throughout the semester is summarized in Table 7. Within this table, the refined interest scores are calculated in such a way that 0 represents the average interest for all students, negative values represent below average interest, and positive values represent above average interest. The $X\beta$ and $\rho Wy + X\beta$ columns are predicting the refined interest score as was done in the modeling steps. At the beginning of the semester, before there was an opportunity for much influence to occur, Alice and Beth both reported high levels of interest while Charlie reported a low level of interest. At the end of the semester, we see that Alice and Beth largely maintained their interest while Charlie grew in

1014 | *Chem. Educ. Res. Pract.*, 2023, **24**, 1003–1024

This journal is © The Royal Society of Chemistry 2023

Table 7 Summary of a group from the data to help illustrate the concept of network effects. Coarse scores are the unweighted means from items in a factor while refined scores are computed from the measurement model. "Model 3 predicted influence without social influence" is the calculation of interest at time 2 ($t2$) from autonomy support ($t2$), perceived competence ($t2$), and relatedness ($t2$) and interest at time 1, without any social influence terms. "Model 3 predicted influence with social influence" includes these same four predictors plus the network effect term (ρWy), though not the network disturbance term

Student	Coarse interest ($t1$)	Coarse interest ($t2$)	Refined interest ($t1$)	Refined interest ($t2$)	Model 3 predicted interest (without social influence)	Model 3 predicted interest (with network effects)
Alice	6.2	6.4	2.080	2.067	1.588	1.576
Beth	6.2	6.0	1.965	1.815	1.674	1.656
Charlie	3.2	2.6	−0.613	−1.256	−0.440	−0.526
Range	3.0	3.8	2.693	3.323	2.114	2.183

his disinterest, expanding the range of interest for this group. When looking at the other terms in the multiple regression model, we see that this drop in Charlie's interest goes beyond what would be expected based on the other variables. For someone with Charlie's initial interest, perception of autonomy support, perceived competence, and feelings of relatedness, we would expect a final interest of around −0.440 in isolation instead of the observed −1.256. The negative values suggests that Charlie has a below average interest with the observed value being even farther below average than the modeled value. Adding the network effect term into the model brings Charlie's expected interest to −0.526 which is still above the observed value but helps reduce the residuals in the model and improves the fit.

Social influence and amount of interaction (research question 4)

Our fourth and final research question was concerned with how the degree of interaction among peers relates to the strength of the social influence. To answer this question, we reran Model 3, but adjusted the networks and students included in the analysis. Specifically, we varied the threshold for the amount of times a pair of students needed to interact to include them as a pair in the analysis. From this data shown in Fig. 3, we see that there is a divide between low and high weight threshold. At low thresholds, we do not see a significant network effect while we do observe significance at higher thresholds. When looking at networks with low weight thresholds, the analysis includes students who perhaps only interacted a single time or two based on groups formed when the peer leaders temporarily combined groups to make up for some students being absent or they might include students who did not consistently attend PLTL sessions. As the students at this level of analysis did not necessarily have a meaningful amount of interaction with each other, they also did not have as many opportunities to influence each other. However, students who frequently interacted with each other (high threshold values) did have the ability to observe each other regularly and had a better ability to judge their relative interests in the course content.

In interpreting our findings of social influence, we also considered that this finding is observed in a largely online environment. While we lack comparable data from previous semesters that would allow us to directly address the potential effects of COVID-19 or online learning, we can make some

reasonable connections to other reported findings. For example, Jeffery and Bauer (2020) found that while lectures may not have changed much for students, the ability to focus on the online content relative to in-person was reduced, which could be a factor in play for the students we have analyzed. Additionally, they also reported that the modal value of exchanges with a student's peer network was reduced by around 90% during this time. Considering our finding that the observed social influence related to the amount of interaction we could quantify from attendance data, a significant change in how much students interact would be expected to affect how much they influence each other. Furthermore, an analysis of collaborative learning in a chemistry setting (Gemmel et al., 2020) showed that, compared to in-person learning, students in the online setting would often take longer to solve problems due to challenges in communication. Additionally, groups that included students with cameras off tended to work less collaboratively. This lack of collaboration among the group members could have had an effect on the type and amount of social influence observed.

Limitations

As is the case for many studies, these data were collected at a single institution over a single semester so care should be taken before generalizing any findings outside the particular place and time being studied. Additionally, these findings are largely related to social influence and interaction which was unique in this semester due to the particular approach to COVID-19 with mixed in-person and online learning. So further studies would be needed to address whether these results might be an artifact of the particular location and time the data were collected or if the findings extend to fully in-person or fully online settings.

Another caution in interpreting these results is that we do not have a causal research design. Any findings should be considered correlative, and any implications of a causal direction are derived from theory and not directly from the data.

Additionally, while our online instruments allowed us to efficiently collect data from many students, we lack some of the nuance available from qualitative data collected by more time intensive means such as interviews with students. Another consideration is that our particular measure for motivation lacks some nuance other researchers have used that considers different types of motivation (e.g. intrinsic and extrinsic) and different elements within each of those types of motivation. We chose our particular measure as it best fit our research

questions and setting, but that decision does limit our ability to speak to how any specific type of motivation is being socially influenced so our claims about motivation are currently quite broad. Another consideration in interpreting the scores from our instrument is that the fit statistics of our CFA did not quite reach the conventionally accepted guidelines. While we feel the level of fit we observed is adequate for the claims we have made here, future research using this instrument should consider that continued development work is probably necessary.

Another limitation in interpreting our result comes from the challenges intrinsic to any form of SNA analysis. As SNA requires a high response rate for reliable analysis (Grosser and Borgatti, 2013), we are limited in our ability to speak to the social influence in any group where we feel we are lacking data, which is part of why our models are run on a relatively small portion of the total student body. While we presented evidence there was not a meaningful difference between the students we analyzed and those we did not (Appendix F), we cannot rule out the possibility that our results only apply to the particular students included in the model and not to the students who were excluded.

Implications for teaching

This research has implications for those teaching in settings where group work is a key focus of the pedagogy which includes many forms of active learning beyond PLTL. As we observe students becoming polarized in their interest in the course activities, it is important to consider how this might harm students who are already entering the course with lower motivation. Additionally, instructors should be cautious if they expect student interest in chemistry for their particular setting might be strongly influenced by demographic categories. One piece of advice that has been given for forming learning teams in gateway courses is to avoid creating situations in which a student from an "at-risk" group is isolated on a team (Felder and Brent, 2001; Oakley *et al.*, 2004). An important idea to accept is that comparisons will happen so it is better to optimize the comparisons than to eliminate them. There are a few potential ways to consider optimizing these comparisons. One potential model for pedagogy is status treatments as laid out by Cohen (1994); Cohen and Lotan (1995). Within status treatments, it is considered important to publicly assign positive evaluation to all students with specific examples of what each student did particularly well. By praising all students, you can help reduce the likelihood of students making upward comparisons and feeling that they are unable to achieve a particular level of success. It is also important that praise be specific because it will otherwise be dismissed. Another consideration is that it is important that the praise be public as written private feedback was not found to have the same positive impact for students (Webster *et al.*, 2003).

Another potential implication can be in the particular applications of a concept an instructor chooses to highlight in class. If all examples used in instruction come from a particular topic (*e.g.* pharmaceuticals), then students interested in the particular topic will consistently grow in interest and this growth may have a continued negative impact on their peers

due to the social comparisons. Therefore, we encourage instructors to consider using a variety of applications to appeal to as many students as possible. For a couple examples, many concepts in general chemistry have been connected to activities such as cooking (Miles and Bachman, 2009; Howell *et al.* 2021) or analyzing the pigments used in paintings (Nivens *et al.*, 2010; Vyhnal *et al.*, 2020). By using a variety of examples, instructors limit the possibility that students will contrast themselves to their peers in such a way that consistently reduces their interest in chemistry.

Implications for research

For researchers, this report highlights how considering the relational nature of data can expand how it is interpreted. By using the tools of SNA, researchers are not limited to considering students as isolated units but can in fact consider them as interconnected. The specific analysis in the paper shows how a multiple regression can be expanded to include social influence terms to consider how students might be affecting each other. Within this research, the social influence was found to be significant and expanded how the data was interpreted. Using this SNA approach is not limited to PLTL or these particular operationalizations of BNT. Many active learning techniques and traditional laboratory courses involve students interacting consistently throughout the semester so treating each student as an isolated unit limits the potential findings. Therefore, we encourage other researchers to apply these tools to a variety of contexts beyond motivation in many other settings than the particular one we observed here.

Conclusions

We performed this research to look into student motivation in a PLTL setting while considering the relational nature of the students. Using multiple regression that did not consider the students' social networks, we found that BNT was a predictive model for student interest in PLTL activities with perceived competence and relatedness consistently predicting student interest. However, mixed support was found for the role of autonomy support in this particular setting. After that analysis, we added a relational component to our analysis through the use of a social influence model. Through the observation of a significant negative network effect process, we found that groups of students had a tendency to polarize in their interest controlling for other factors. This finding seems consistent with a contrast effect described by social comparison theory where through comparisons with their peers, students become more entrenched in their relative disposition toward chemistry. By adjusting our analysis based on the amount of times a pair of students interacted, we observed that social influence required a certain threshold of interaction before the influence was statistically observable which was 8 interactions in our data. These findings presented here suggest that instructors ought to consider how to appeal to the interests of a variety of students so as to not exacerbate preexisting differences. Beyond

1016 | *Chem. Educ. Res. Pract.*, 2023, **24**, 1003–1024

This journal is © The Royal Society of Chemistry 2023

the particular findings for this setting and study, this work also demonstrates a method for researchers in modeling relational data to better consider how students are affecting each other in their outcomes measures.

Conflicts of interest

There are no conflicts of interest to declare.

Appendices

Appendix A: instrument items

Students were first presented with items adapted from the Learning Climate Questionnaire (LCQ) to characterize autonomy support (Williams and Deci, 1996). Items answered on a 7-point scale ranging from Strongly Disagree to Strongly Agree. Text that originally appeared as "my instructor" was replaced with "my peer leader" to fit the setting and research questions. Item number 13 is reverse phrased and was reverse coded for ease of interpretation.

LCQ items:

1. I feel that my peer leader provides me choices and options.
2. I feel understood by my peer leader.
3. I am able to be open with my peer leader during class.
4. My peer leader conveyed confidence in my ability to do well in this course.
5. I feel that my peer leader accepts me.
6. My peer leader made sure I really understood the goals of the course and what I need to do.
7. My peer leader encouraged me to ask questions.
8. I feel a lot of trust in my peer leader.
9. My peer leader answers my questions fully and carefully.
10. My peer leader listens to how I would like to do things.
11. My peer leader handles people's emotions very well.
12. I feel that my peer leader cares about me as a person.
13. I don't feel very good about the way my peer leader talks to me.
14. My peer leader tries to understand how I see things before suggesting a new way to do things.
15. I feel able to share my feelings with my peer leader.

Students were then presented with items adapted from the Intrinsic Motivation Inventory (IMI) to characterize interest, perceived competence, and relatedness (McAuley *et al.*, 1989). Items were answered on a 7-point scale ranging from "Not true at all" to "Very true" In adapting items from the original source, "this activity" was replaced with "Peer Leading Activities" and subject/verb agreement was adjusted as necessary. For items characterizing relatedness, "this person" became "my peer leading group" to match the context. Items 2, 3, 6, 11, 18, 20, and 21 are reverse phrased and were reverse coded for ease of interpretation.

The items break down by intended factor in the following way:

- Interest: 1, 4, 9, 11, 13, 17, 20
- Perceived competence: 3, 5, 7, 10, 15, 19
- Relatedness: 2, 6, 8, 12, 14, 16, 18, 21

IMI items:

1. I would describe Peer Leading Activities as very interesting.
2. I don't feel like I could really trust my peer leading group.
3. The Peer Leading Activities were ones that I couldn't do very well.
4. I thought Peer Leading Activities were quite enjoyable.
5. After working at Peer Leading Activities for awhile, I felt pretty competent.
6. I felt really distant to my peer leading group.
7. I think I am pretty good at Peer Leading Activities.
8. It is likely that my peer leading group and I could become friends if we interacted a lot.
9. Peer Leading Activities were fun to do.
10. I am satisfied with my performance at Peer Leading Activities.
11. Peer Leading Activities did not hold my attention at all.
12. I feel close to my peer leading group.
13. While I was doing Peer Leading Activities, I was thinking about how much I enjoyed them.
14. I'd like a chance to interact with my peer leading group more often.
15. I think I did pretty well at Peer Leading Activities, compared to other students.
16. I felt like I could really trust my peer leading group.
17. I enjoyed doing Peer Leading Activities very much.
18. I really doubt that my peer leading group and I would ever be friends.
19. I was pretty skilled at Peer Leading Activities.
20. I thought Peer Leading Activities were boring.
21. I'd really prefer not to interact with my peer leading group in the future.

Appendix B: response rate breakdown

As it was possible that those who completed the instrument and consented to participating in our study might not represent the class population, we chose to analyze the rate of participation in each subgroup analyzed. The summary of these rates is presented in Table 8 below. These data helped us determine if a group was over or under-represented in our sampled data. We did not see evidence of that occurring.

Table 8 Number and percentage of students who completed our instrument and consented to participate in our research study at each time point

	n	$t1$	$t2$	$t1$ and $t2$
All students	1988	1179 (59%)	1244 (63%)	870 (44%)
Female	1167	759 (65%)	762 (65%)	569 (49%)
Male	821	420 (51%)	482 (59%)	301 (37%)
White	784	475 (61%)	507 (65%)	353 (45%)
Hispanic	428	261 (61%)	270 (63%)	201 (47%)
First time in college	1711	1037 (61%)	1096 (64%)	772 (45%)
Transfer students	220	107 (49%)	113 (51%)	75 (34%)

This journal is © The Royal Society of Chemistry 2023

Chem. Educ. Res. Pract., 2023, **24**, 1003–1024 | **1017**

Appendix C: exploratory factor analysis

Initial assumption checking

Before conducting factor analysis, we checked the data against the assumptions utilized in the method (*e.g.* normality). As part of this, we produced Table 9 which summarizes the mean, standard deviation, skewness, and kurtosis for each item. Generally, factor analysis is well behaved if the absolute skewness and kurtosis are less than 2 though some support a more liberal absolute kurtosis that can go all the way to 7 (Bandalos and Finney, 2018). For ease of interpretation, each item is named for the instrument it came from, followed by the item number. For the IMI, ".int", ".pc", ".rel" was added to identify the item as intended for interest, perceived competence, and relatedness respectively. Finally, a ".r" was added to all items which were presented as reverse phrased.

Additionally, we checked the full inter-item correlation table (630 values), and nothing from that was flagged as problematic. We also checked the frequency that each item was left blank by a student and found that no item had more than 0.5% missing data.

Determining number of factors

As mentioned in the Methods section, data were split in half before performing EFA and CFA using a random sort, and this was performed for each administration of the instrument. To determine the number of factors that underlie the data for EFA, we utilized a combination of techniques including theoretical considerations, Kaiser's criterion, scree analysis, and parallel analysis. The *nFactors* library version 2.4.1 in R (Raiche and Magis, 2020) was utilized for this analysis. From these analyses, we saw evidence that the instrument was divisible into 4 (theoretical), 5 (scree and parallel analysis), or 6 (Kaiser's criterion) factors.

Full instrument EFA

Factor structures were then calculated at each time point based on all reasonable number of factors. This analysis utilized the *psych* library version 2.1.9 in R (Revelle, 2021). As it would be overwhelming to present all the EFA results in detail, only results from the models which most informed our thinking will be provided. Readers interested in viewing any other EFA outputs that were part of our analysis can contact the corresponding author. The presented analyses were performed on data collected in the second round of data collection.

Table 9 Descriptive analysis of received data by item

Item	Mean	SD	Skewness	Kurtosis
lcq.01	5.60	1.30	−1.22	4.48
lcq.02	5.67	1.23	−1.11	4.26
lcq.03	5.60	1.30	−1.08	4.00
lcq.04	5.71	1.24	−1.15	4.36
lcq.05	5.96	1.08	−1.24	4.82
lcq.06	5.80	1.22	−1.29	4.76
lcq.07	6.02	1.12	−1.46	5.46
lcq.08	5.52	1.31	−0.88	3.57
lcq.09	5.99	1.09	−1.43	5.66
lcq.10	5.41	1.32	−0.64	2.87
lcq.11	5.47	1.22	−0.47	2.61
lcq.12	5.35	1.27	−0.46	2.75
lcq.13.r	5.93	1.50	−1.75	5.36
lcq.14	5.39	1.23	−0.50	2.82
lcq.15	4.94	1.43	−0.37	2.70
imi.01.int	4.44	1.69	−0.26	2.50
imi.02.rel.r	5.53	1.61	−1.07	3.46
imi.03.pc.r	5.35	1.63	−0.83	2.94
imi.04.int	4.31	1.67	−0.15	2.51
imi.05.pc	5.04	1.54	−0.48	2.75
imi.06.rel.r	4.68	1.88	−0.42	2.17
imi.07.pc	4.88	1.56	−0.32	2.60
imi.08.rel	4.11	1.72	−0.01	2.37
imi.09.int	4.10	1.68	0.03	2.44
imi.10.pc	5.05	1.61	−0.56	2.74
imi.11.int.r	5.04	1.72	−0.69	2.68
imi.12.rel	2.91	1.66	0.64	2.76
imi.13.int	3.00	1.73	0.62	2.67
imi.14.rel	3.72	1.76	0.19	2.34
imi.15.pc	4.35	1.60	−0.11	2.61
imi.16.rel	4.03	1.65	0.03	2.50
imi.17.int	3.90	1.72	0.11	2.38
imi.18.rel.r	4.68	1.76	−0.44	2.39
imi.19.pc	4.54	1.56	−0.14	2.63
imi.20.int.r	4.62	1.77	−0.42	2.40
imi.21.rel.r	5.34	1.71	−0.98	3.22

Table 10 Factor loadings from 4-factor EFA with all items. Values below 0.3 are suppressed for improved readability

	Autonomy support	Combined interest and relatedness	Perceived competence	Reverse phrased items
lcq.01	0.69			
lcq.02	0.79			
lcq.03	0.77			
lcq.04	0.79			
lcq.05	0.81			
lcq.06	0.80			
lcq.07	0.76			
lcq.08	0.83			
lcq.09	0.77			
lcq.10	0.79			
lcq.11	0.75			
lcq.12	0.77			
lcq.14	0.71			
lcq.15	0.68	0.34		
imi.01.int	0.38	0.56		
imi.04.int		0.60	0.41	
imi.09.int		0.70	0.40	
imi.13.int		0.75	0.27	
imi.17.int		0.74	0.34	
imi.08.rel		0.60		
imi.12.rel		0.74		
imi.14.rel		0.66		
imi.16.rel	0.34	0.60		
imi.05.pc	0.35	0.31	0.55	
imi.07.pc			0.83	
imi.10.pc			0.73	
imi.15.pc			0.71	
imi.19.pc			0.79	
lcq.13.r				0.37
imi.11.int.r				0.67
imi.20.int.r	0.32			0.62
imi.02.rel.r				0.67
imi.06.rel.r				0.62
imi.18.rel.r				0.65
imi.21.rel.r				0.69
imi.03.pc.r			0.43	0.41

Chemistry Education Research and Practice

Results from running a 4-factor EFA on all items are presented in Table 10. In this analysis, we observed that reverse phrased items seemed to load onto a single factor together regardless of the intended factor for each item and this behavior was also observed in a 3-factor solution and at both administrations. For ease of interpretation for the instrument scores, we eliminated the reverse phrased items from our analysis.

After removal of reverse phrased items

After removing the reverse phrased items, we repeated the analysis of determining number of factors in the remaining items. At both time points, we saw evidence of a 3-factor solution (parallel analysis) or 4-factor solution (theoretical considerations, scree analysis, and Kaiser's criterion).

Factor structures were again calculated based on both potential number of factors. The summary of the item loadings can be found in Table 11 for a 3-factor solution and Table 12 for a 4-factor solution. From these results as well as the consideration in determining number of potential factors, we could reasonably propose two different factor structures. One was a 4-factor solution which directly matches the intended design of the instrument (autonomy support, interest, perceived competence, and relatedness). However, we could not rule out a separate 3-factor structure which divided the items into factors of autonomy support, perceived competence, and a factor that combined items from interest and relatedness together.

Table 11 Factor loadings from 3-factor EFA without reverse phrased items. Values below 0.3 are suppressed for improved readability

	Autonomy support	Perceived competence	Combined relatedness and interest
lcq.01	0.64		
lcq.02	0.79		
lcq.03	0.76		
lcq.04	0.74		
lcq.05	0.74		
lcq.06	0.78		
lcq.07	0.71		
lcq.08	0.82		
lcq.09	0.74		
lcq.10	0.75		
lcq.11	0.74		
lcq.12	0.71		
lcq.14	0.67		
lcq.15	0.70		
imi.05.pc	0.34	0.59	0.32
imi.07.pc		0.81	
imi.10.pc		0.74	
imi.15.pc		0.75	
imi.19.pc		0.81	
imi.08.rel			0.58
imi.12.rel			0.70
imi.14.rel			0.61
imi.16.rel	0.35		0.60
imi.01.int	0.37		0.57
imi.04.int		0.43	0.65
imi.09.int		0.38	0.73
imi.13.int			0.73
imi.17.int		0.36	0.75

Table 12 Factor loadings from 4-factor EFA without reverse phrased items. Values below 0.3 are suppressed for improved readability

	Autonomy support	Perceived competence	Relatedness	Interest
lcq.01	0.64			
lcq.02	0.79			
lcq.03	0.76			
lcq.04	0.74			
lcq.05	0.74			
lcq.06	0.77			
lcq.07	0.70			
lcq.08	0.81			
lcq.09	0.74			
lcq.10	0.75			
lcq.11	0.74			
lcq.12	0.72			
lcq.14	0.67			
lcq.15	0.71		0.33	
imi.05.pc	0.34	0.56		0.33
imi.07.pc		0.81		
imi.10.pc		0.73		
imi.15.pc		0.80		
imi.19.pc		0.84		
imi.08.rel			0.61	
imi.12.rel			0.73	
imi.14.rel			0.63	
imi.16.rel	0.35		0.62	
imi.01.int		0.33	0.53	0.60
imi.04.int	0.36		0.36	0.59
imi.09.int		0.33	0.57	0.53
imi.13.int	0.37		0.41	0.67
imi.17.int			0.62	0.38

Therefore, we used CFA to compare these models to find the one that was a better approximation for our data.

Appendix D: confirmatory factor analysis

CFA was used to compare a potential 3-factor solution and 4-factor solution to our data. CFA models were run using the *lavaan* library version 0.6.9 in R (Rosseel, 2012) with a full-information maximum likelihood estimator. The fit indices for these models are summarized in Tables 13 and 14. As is often the case with cutoff values in statistics, fit statistic cutoff values are somewhat arbitrary and may not even be appropriate for some data sets. For example, the conventional Hu and Bentler (1999) guidelines were developed based on a relatively narrow range of loadings (0.7–0.8) and going below (Heene *et al.*, 2011) or above (Browne *et al.*, 2002) this range can affect the chance of accepting or rejecting a model based on the standard guidelines. We will still reference the conventional guidelines when discussing our fit statistics; however, we will not use them as firm cutoffs.

At both time points, we see evidence that there is a substantial improvement of fit by incorporating a fourth factor into the model. While some level of fit improvement is to be expected by adding any additional parameters, it is noteworthy that the improvement to χ^2 is statistically significant. Additionally, a parsimony adjusted fit index (RMSEA), sees improvement even with a less parsimonious model. Therefore, when

Table 13 Comparison of 3 and 4-factor solutions to instrument at first administration

Number factors	χ^2	df	p-Value	SRMR	RMSEA	CFI	$\Delta\chi^2$	Δdf	p-Value	ΔSRMR	ΔRMSEA	ΔCFI
4	1699	344	<0.001	0.058	0.082	0.888	—	—	—	—	—	—
3	1986	347	<0.001	0.064	0.089	0.865	289	3	<0.001	0.006	0.007	−0.023

Table 14 Comparison of 3 and 4-factor solutions to instrument at second administration

Number factors	χ^2	df	p-Value	SRMR	RMSEA	CFI	$\Delta\chi^2$	Δdf	p-Value	ΔSRMR	ΔRMSEA	ΔCFI
4	1692	344	<0.001	0.064	0.079	0.896	—	—	—	—	—	—
3	1919	347	<0.001	0.067	0.085	0.878	226	3	<0.001	0.003	0.006	−0.018

considering these data and the theoretical design of the instrument, we believe the 4-factor solution provides a better approximation of our data.

The 4-factor solution calculated at the second administration is illustrated in Fig. 4. When looking at the model fit indices for the 4-factor solution, we see that χ^2 indicates significant misfit. For the other fit indices, we see that SRMR is within the generally accepted guidelines (Hu and Bentler, 1999) for good levels of fit (<0.08); however, we find that RMSEA and CFI fall outside the traditional guidelines (RMSEA < 0.06; CFI > 0.95). There are some sources

Table 15 Fit statistics for comparison of possible alternative measurement models

	χ^2	df	SRMR	RMSEA	CFI
Model used (reverse items dropped)	1692	344	0.064	0.079	0.896
All items	3251	588	0.075	0.085	0.830
With negative 'method' factor	2392	577	0.065	0.071	0.884
Bifactor	2581	558	0.098	0.076	0.871

that suggests RMSEA < 0.08 (Bandalos and Finney, 2018) and CFI > 0.9 (McDonald and Ho, 2002) can be considered acceptable levels of fit. For interpretability, we elected to retain the 4-factor model.

At this point in the CFA, we chose to explore alternate measurement models that utilized more of our collected data,. For this task, we first ran a CFA that included all items based on their intended theoretical scales. The next CFA we conducted incorporated a negative 'methods' factor which is a technique that has been shown to improve model fit (Naibert and Barbera, 2022) in cases similar to ours. Finally, we ran a bifactor model as that might also have provided a way to better understand our data. The fit statistics for these three models as well as the fit statistics for the model shown in Fig. 4 are shown in Table 15. Since the fit of none of these models represented a meaningful improvement over the model in Fig. 4, we elected to retain the four-factor model without reverse phrased items as this model is the most straightforward to interpret.

Appendix E: measurement invariance testing

We used measurement invariance testing to provide evidence that it is appropriate to make comparisons between the instrument given at different times and to different populations.

R code to calculate measurement invariance was adapted from Rocabado *et al.* (2020) and used the *lavaan* library (Rosseel, 2012). According to guidelines for invariance laid out by Chen (2007), it is acceptable to move up the steps of measurement invariance testing if the fit indices do not become worse by certain thresholds. Generally, we are looking for ΔSRMR > 0.03, ΔRMSEA > 0.015, and ΔCFI < −0.01 as
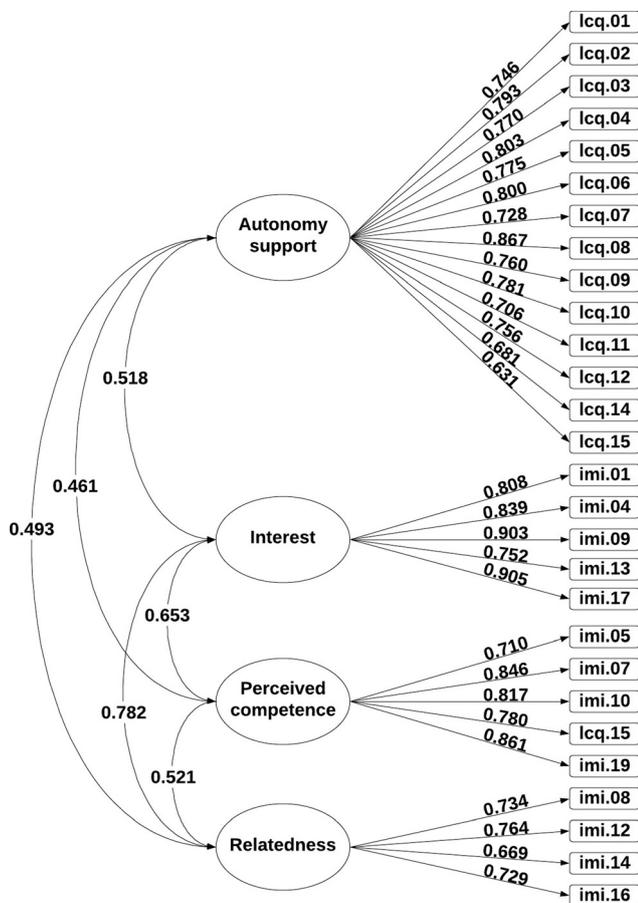

Fig. 4 Standardized CFA showing the results of the 4-factor solution calculated from the second administration of our instrument.

establishing a lack of invariance. For establishing scalar or strict invariance, the threshold for ΔSRMR is brought down to 0.01. Additionally, Chen laid out stricter guidelines (ΔSRMR $>$ 0.025, ΔRMSEA $>$ 0.01, and ΔCFI < -0.005 for metric invariance with ΔSRMR > 0.005 for scalar and strict invariance) when the size of groups is small or unevenly distributed. As our sample had a limited number of transfer and Hispanic students ($n < 300$), we aimed for the stricter guidelines on these models.

We ran 4 sets of measurement invariance tests to determine if we could fairly compare our data. These comparisons considered whether measurement invariance occurred between the two separate administrations (Table 16), non-transfer and transfer students (Table 17), female and male students (Table 18), and White and Hispanic students (Table 19). In all four cases, we accepted that we met standards for strict invariance allowing us to use factor scores to directly compare the investigated groups.

Appendix F: comparison between included and excluded students

To examine social influence in our models, we had to exclude students who did not complete both instruments and were not paired with at least 2 other students who also completed both instruments. It is possible that the 270 students included are not representative of the class so we investigated how the included students differed from their peers. The tables below compare the coarse (Table 20) and refined (Table 21) factor scores between the students who were included in our models and those from whom we had factor scores but were not among those modeled. Then Table 22 shows a summary of how the demographics of students included in the model compare to the rest of the students in the class. From this data, we did not see strong evidence of difference though it may be that the

Table 16 Summary of measurement invariance between time 1 (n = 1179) and time 2 (n = 1244)

Step	Testing level	χ^2	df	p-Value	SRMR	RMSEA	CFI	$\Delta\chi^2$	Δdf	p-Value	ΔSRMR	ΔRMSEA	ΔCFI
0	Baseline (time 1)	2761	344	<0.001	0.056	0.077	0.899	—	—	—	—	—	—
0	Baseline (time 2)	2705	344	<0.001	0.060	0.074	0.908	—	—	—	—	—	—
1	Configural	5466	688	<0.001	0.058	0.076	0.904	—	—	—	—	—	—
2	Metric	5535	712	<0.001	0.061	0.075	0.903	69	24	<0.001	0.003	−0.001	−0.001
3	Scalar	5643	736	<0.001	0.062	0.074	0.901	107	24	<0.001	0.001	−0.001	−0.002
4	Strict	5762	764	<0.001	0.062	0.073	0.899	119	28	<0.001	<0.001	−0.001	−0.002

Table 17 Summary of measurement invariance between transfer (n = 113) *vs.* non-transfer (n = 1096) students

Step	Testing level	χ^2	df	p-Value	SRMR	RMSEA	CFI	$\Delta\chi^2$	Δdf	p-Value	ΔSRMR	ΔRMSEA	ΔCFI
0	Baseline (transfer)	703	344	<0.001	0.082	0.096	0.858	—	—	—	—	—	—
0	Baseline (non-transfer)	2498	344	<0.001	0.061	0.074	0.908	—	—	—	—	—	—
1	Configural	3202	688	<0.001	0.063	0.077	0.903	—	—	—	—	—	—
2	Metric	3234	712	<0.001	0.064	0.075	0.903	32	24	<0.001	0.001	−0.002	<0.001
3	Scalar	3275	736	<0.001	0.064	0.074	0.902	40	24	<0.001	<0.001	−0.001	−0.001
4	Strict	3323	764	<0.001	0.065	0.073	0.901	48	28	<0.001	0.001	−0.001	−0.001

Table 18 Summary of measurement invariance between female (n = 762) and male (n = 482) students

Step	Testing level	χ^2	df	p-Value	SRMR	RMSEA	CFI	$\Delta\chi^2$	Δdf	p-Value	ΔSRMR	ΔRMSEA	ΔCFI
0	Baseline (Female)	1795	344	<0.001	0.058	0.074	0.913	—	—	—	—	—	—
0	Baseline (Male)	1374	344	<0.001	0.068	0.079	0.888	—	—	—	—	—	—
1	Configural	3170	688	<0.001	0.062	0.076	0.904	—	—	—	—	—	—
2	Metric	3198	712	<0.001	0.063	0.075	0.904	28	24	0.240	0.001	−0.001	<0.001
3	Scalar	3264	736	<0.001	0.065	0.074	0.902	65	24	<0.001	0.002	−0.001	−0.002
4	Strict	3342	764	<0.001	0.065	0.074	0.900	78	28	<0.001	<0.001	<0.001	−0.002

Table 19 Summary of measurement invariance between White (n = 507) *vs.* Hispanic (n = 270) students

Step	Testing level	χ^2	df	p-Value	SRMR	RMSEA	CFI	$\Delta\chi^2$	Δdf	p-Value	ΔSRMR	ΔRMSEA	ΔCFI
0	Baseline (White)	1500	344	<0.001	0.069	0.081	0.897	—	—	—	—	—	—
0	Baseline (Hispanic)	967	344	<0.001	0.071	0.082	0.878	—	—	—	—	—	—
1	Configural	2467	688	<0.001	0.070	0.082	0.891	—	—	—	—	—	—
2	Metric	2497	712	<0.001	0.072	0.080	0.891	29	24	0.209	0.002	−0.002	<0.001
3	Scalar	2514	736	<0.001	0.072	0.079	0.892	17	24	0.825	<0.001	−0.001	0.001
4	Strict	2624	764	<0.001	0.073	0.079	0.887	109	28	<0.001	0.001	<0.001	−0.005

Table 20 Comparison of coarse factor scores between students included in statistical models and those not included

	Modeled	Unmodeled	difference
Interest ($t1$)	4.04	4.21	−4%
Interest ($t2$)	3.78	3.72	2%
Autonomy support ($t1$)	5.44	5.60	−3%
Autonomy support ($t2$)	5.60	5.65	−1%
Perceived competence ($t1$)	4.72	4.83	−2%
Perceived competence ($t2$)	4.76	4.73	1%
Relatedness ($t1$)	5.30	5.49	−3%
Relatedness ($t2$)	3.54	3.46	2%

Table 21 Comparison of refined factor scores between students included in statistical models and those not included. Difference should be understood as 1 being around a standard deviation from the typical response (0)

	Modeled	Unmodeled	Difference
Interest ($t1$)	−0.14	0.04	−0.18
Interest ($t2$)	−0.43	−0.50	0.06
Autonomy support ($t1$)	−0.11	0.03	−0.15
Autonomy support ($t2$)	0.04	0.08	−0.04
Perceived competence ($t1$)	−0.08	0.02	−0.11
Perceived competence ($t2$)	−0.04	−0.06	0.02
Relatedness ($t1$)	−0.09	0.03	−0.11
Relatedness ($t2$)	−0.37	−0.43	0.06

Table 22 Comparison of demographics of interest between students included in statistical models and those not included

	Modeled	Unmodeled	Difference
Female	67%	60%	7%
Male	33%	40%	−7%
White	41%	40%	1%
Hispanic	23%	21%	2%
Non-transfer	91%	87%	4%
Transfer	7%	10%	−3%

computed models are more representative of female students than male students.

Acknowledgements

References

American Education Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA & NCME), (2014), *Standards for educational and psychological testing*, American Education Research Association, 4th edn, Washington DC: AERA Publication Sales.

Arjoon J. A., Xu X. and Lewis J. E., (2013), Understanding the state of the art for measurement in chemistry education research: examining the psychometric evidence, *J. Chem. Educ.*, **90**(5), 536–545.

Bandalos D. L. and Finney S. J., (2018), Factor analysis, in Hancock G. R., Stapleton L. M. and Mueller R. O. (ed.), *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, New York, NY: Routledge, pp. 98–122.

Baumeister R. F. and Leary M. R., (1995), The need to belong: desire for interpersonal attachments as a fundamental human motivation, *Psychol. Bull.*, **117**(3), 57–89.

Berg S. A. and Moon A., (2022), Prompting hypothetical social comparisons to support chemistry students' data analysis and interpretations, *Chem. Educ. Res. Pract.*, **23**(1), 124–136.

Black A. E. and Deci E. L., (2000), The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: a self-determination theory perspective, *Sci. Educ.*, **84**(6), 740–756.

Bongers A., (2022), Virtual poster session designed for social cognitive learning in undergraduate chemistry research, *J. Chem. Educ.*, **99**(6), 2259–2269.

Brewe E., Kramer L. and Sawtelle V., (2012), Investigating student communities with network analysis of interactions in a physics learning center, *Phys. Rev. ST Phys. Educ. Res.*, **8**(1), 010101.

Brown T. A., (2015), The common factor model and exploratory factor analysis, in *Confirmatory factor analysis for applied research*, New York, NY: Guilford Publications, pp. 10–32.

Browne M. W., MacCallum R. C., Kim C. T., Andersen B. L. and Glaser R., (2002), When fit indices and residuals are incompatible, *Psychol. Methods*, **7**(4), 403.

Butts C. T., (2020), sna: Tools for Social Network Analysis, *R Package version 2.6*.

Chan J. Y. and Bauer C. F., (2015), Effect of peer-led team learning (PLTL) on student achievement, attitude, and self-concept in college general chemistry in randomized and quasi experimental designs, *J. Res. Sci. Teach.*, **52**(3), 319–346.

Chen F. F., (2007), Sensitivity of goodness of fit indexes to lack of measurement invariance, *Struc. Equ. Modeling*, **14**(3), 464–504.

Cicuto C. A. T. and Torres B. B., (2016), Implementing an active learning environment to influence students' motivation in biochemistry, *J. Chem. Educ.*, **93**(6), 1020–1026.

Cohen J., (1988), The *t* test for means. in *Statistical power analysis for the behavioral sciences*, New York: Routledge, pp. 19–74.

Cohen E. G., (1994), Restructuring the classroom: conditions for productive small groups, *Rev. Educ. Res.*, **64**(1), 1–35.

Cohen E. G. and Lotan R. A., (1995), Producing equal-status interaction in the heterogeneous classroom, *Am. Educ. Res. J.*, **32**(1), 99–120.

1022 | *Chem. Educ. Res. Pract.*, 2023, **24**, 1003–1024

This journal is © The Royal Society of Chemistry 2023

Deci E. L. and Ryan R. M., (1985), An Introduction, in *Intrinsic motivation and self-determination in human behavior*, New York, NY: Plenum Press, pp. 3–10.

Deci E. L. and Vansteenkiste M., (2004), Self-determination theory and basic need satisfaction: understanding human development in positive psychology, *Ricerche di Psicologia*, **27**(1), 23–40.

Dijkstra P., Kuyper H., Van der Werf, G., Buunk, A. P. and van der Zee, Y. G., (2008), Social comparison in the classroom: a review, *Rev. Educ. Res.*, **78**(4), 828–879.

Dou R., Brewe E., Zwolak J. P., Potvin G., Williams E. A. and Kramer L. H., (2016), Beyond performance metrics: examining a decrease in students' physics self-efficacy through a social networks lens, *Phys. Rev. Phys. Educ. Res.*, **12**(2), 020124.

Felder R. M. and Brent R., (2001), Effective strategies for cooperative learning, *J. Cooperation Collaboration College Teach.*, **10**(2), 69–75.

Felson R. B. and Reed M. D., (1986), Reference groups and self-appraisals of academic ability and performance, *Soc. Psychol. Quart.*, **49**(2), 103–109.

Ferreira D. M., Sentanin F. C., Parra K. N., Negrao Bonini V. M., de Castro M. and Kasseboehmer A. C., (2022), Implementation of inquiry-based science in the classroom and its repercussion on the motivation to learn chemistry, *J. Chem. Educ.*, **99**(2) 578–591.

Ferrell B., Phillips M. M. and Barbera J., (2016), Connecting achievement motivation to performance in general chemistry, *Chem. Educ. Res. Prac.*, **17**(4), 1054–1066.

Festinger L., (1954), A theory of social comparison processes, *Hum. Relat.*, 7(2), 117–140.

Flaherty A. A., (2020), A review of affective chemistry education research and its implications for future research, *Chem. Educ. Res. Pract.*, **21**(3), 698–713.

Frey R. F., Fink A., Cahill M. J., McDaniel M. A. and Solomon E. D., (2018), Peer-led team learning in general chemistry I: interactions with identity, academic preparation, and a course-based intervention, *J. Chem. Educ.*, **95**(12), 2103–2113.

Gemmel P. M., Goetz M. K., James N. M., Jesse K. A., Ratliff B. J., (2020), Collaborative learning in chemistry: impact of COVID-19, *J. Chem. Educ.*, **97**(9), 2899–2904.

Goodwin J. A. and Gilbert B. D., (2001), Cafeteria-style grading in general chemistry, *J. Chem. Educ.*, **78**(4), 490.

Gosser D. K. and Roth V., (1998), The workshop chemistry project: peer-led team-learning, *J. Chem. Educ.*, **75**(2), 185.

Grosser T. J. and Borgatti S. P., (2013), Network theory/social network analysis, in McGee R. J. and Warms R. L. (ed.), *Theory in social and cultural anthropology: An encyclopedia*, Thousand Oaks, CA: SAGE Publications, Inc., pp. 595–597.

Grunspan D. Z., Wiggins B. L. and Goodreau S. M., (2014), Understanding classrooms through social network analysis: a primer for social network analysis in education research, *CBE—Life Sci. Educ.*, **13**(2), 167–178.

Heene M., Hilbert S., Draxler C., Ziegler M. and Bühner M., (2011), Masking misfit in confirmatory factor analysis by increasing unique variances: a cautionary note on the usefulness of cutoff values of fit indices, *Psychol. Methods*, **16**(3), 319.

Hibbard L., Sung S. and Wells B., (2016), Examining the effectiveness of a semi-self-paced flipped learning format in a college general chemistry sequence, *J. Chem. Educ.*, **93**(1), 24–30.

Howell E. L., Yang S., Holesovsky C. M. and Scheufele D. A., (2021), Communicating chemistry through cooking and personal health: everyday applications increase perceived relevance, interest, and self-efficacy in chemistry, *J. Chem. Educ.*, **98**(6), 1852–1862.

Hu L. T. and Bentler P. M., (1999), Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, *Struc. Equ. Modeling*, **6**(1), 1–55.

Jeffery K. A. and Bauer C. F., (2020), Students' responses to emergency remote online teaching reveal critical factors for all teaching, *J. Chem. Educ.*, **97**(9), 2472–2485.

Jewett S., Sutphin K., Gierasch T., Hamilton P., Lilly K., Miller K., Newlin D., Pires R., Sherer M. and LaCourse W. R., (2018), Awareness, analysis, and action: curricular alignment for student success in general chemistry, *J. Chem. Educ.*, **95**(2), 242–247.

Jurišević M., Vrtačnik M., Kwiatkowski M. and Gros N., (2012), The interplay of students' motivational orientations, their chemistry achievements and their perception of learning within the hands-on approach to visible spectrometry, *Chem. Educ. Res. Pract.*, **13**(3), 237–247.

Kaufman J. C., Agars M. D. and Lopez-Wagner M. C., (2008), The role of personality and motivation in predicting early college academic success in non-traditional students at a Hispanic-serving institution, *Learn. Individ. Differ.*, **18**(4), 492–496.

Lane A. K., Skvoretz J., Ziker J. P., Couch B. A., Earl B., Lewis J. E., McAlpin J. D., Prevost L. B., Shadle S. E. and Stains M., (2019), Investigating how faculty social networks and peer influence relate to knowledge and use of evidence-based teaching practices, *Int. J. STEM Educ.*, **6**(1), 1–14.

Leenders R. T. A., (2002), Modeling social influence through network autocorrelation: constructing the weight matrix, *Soc. Networks*, **24**(1), 21–47.

Lewis S. E., (2022), Considerations on validity for studies using quantitative data in chemistry education research and practice, *Chem. Educ. Res. Pract.*, **23**(4), 764–767.

Liu Y., (2017), Investigating Students' Basic Needs and Motivation in College Chemistry Courses with the Lens of Self-Determination Theory (Publication No. 10603525) [Doctoral dissertation, University of South Florida]. Proquest LLC.

Liu Y., Ferrell B., Barbera J. and Lewis J. E., (2017), Development and evaluation of a chemistry-specific version of the academic motivation scale (AMS-Chemistry), *Chem. Educ. Res. Pract.*, **18**(1), 191–213.

Liu Y., Raker J. R. and Lewis J. E., (2018), Evaluating student motivation in organic chemistry courses: moving from a lecture-based to a flipped approach with peer-led team learning, *Chem. Educ. Res. Pract.*, **19**(1), 251–264.

This journal is © The Royal Society of Chemistry 2023

Chem. Educ. Res. Pract., 2023, **24**, 1003–1024 | **1023**

Liyanage D., Lo S. M. and Hunnicutt S. S., (2021), Student discourse networks and instructor facilitation in process oriented guided inquiry physical chemistry classes, *Chem. Educ. Res. Pract.*, **22**(1), 93–104.

Lloyd B. W. and Spencer J. N., (1994), The forum: new directions for general chemistry: recommendations of the task force on the general chemistry curriculum, *J. Chem. Educ.*, **71**(3), 206–209.

Mason D. and Mittag K. C., (2001), Evaluating the success of Hispanic-surname students in first-semester general chemistry, *J. Chem. Educ.*, **78**(2), 256.

McAuley E., Duncan T. and Tammen V. V., (1989), Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis, *Res. Q. Exercise Sport*, **60**(1), 48–58.

McDonald R. P. and Ho M. H. R., (2002), Principles and practice in reporting structural equation analyses, *Psychol Methods*, **7**(1), 64–82.

Miles D. T. and Bachman J. K., (2009), Science of food and cooking: a non-science majors course, *J. Chem. Educ.*, **86**(3), 311.

Naibert N. and Barbera J., (2022), Development of evaluation of a survey to measure student engagement at the activity level in general chemistry, *J. Chem. Educ.*, **99**(3), 1410–1419.

NCES, (2020), IPEDS 12 Month Enrollment Survey. **https://nces.ed.gov/ipeds/use-the-data**.

Nivens D. A., Padgett C. W., Chase J. M., Verges K. J. and Jamieson D. S., (2010), Art, meet chemistry; Chemistry, meet art: case studies, current literature, and instrumental methods combined to create a hands-on experience for nonmajors and instrumental analysis students, *J. Chem. Educ.*, **87**(10), 1089–1093.

Oakley B., Felder R. M., Brent R. and Elhajj I., (2004), Turning student groups into effective teams, *J. Student Centered Learn.*, **2**(1), 9–34.

Ogunde J. C., Overton T. L., Thompson C. D., Mewis R. and Boniface S., (2017), Beyond graduation: motivations and career aspirations of undergraduate chemistry students. *Chem. Educ. Res. Pract.*, **18**(3), 457–471.

Raiche G. and Magis D., (2020), nFactors: parallel analysis and other non graphical solutions to the Cattell scree test, *R package version 2.4.1*.

R Core Team (2021), *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing.

Revelle W., (2021) psych: procedures for personality and psychological research, *R package version 2.1.9*, Evanston, Illinois, USA: Northwestern University.

Rocabado G. A., Kilpatrick N. A., Mooring S. R. and Lewis J. E., (2019), Can we compare attitude scores among diverse populations? An exploration of measurement invariance testing to support valid comparisons between black female students and their peers in an organic chemistry course, *J. Chem. Educ.*, **96**(11), 2371–2382.

Rocabado G. A., Komperda R., Lewis J. E. and Barbera J., (2020), Addressing diversity and inclusion through group comparisons: a primer on measurement invariance testing, *Chem. Educ. Res. Pract.*, **21**(3), 969–988.

Rosseel Y., (2012), lavaan: an R package for structural equation modeling, *J. Stat. Softw.*, **48**(2), 1–36.

Ryan R. M. and Deci E. L., (2000), Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, *Am. Psychol.*, **55**(1), 68.

Scariano S. M. and Davenport J. M., (1987), The effects of violations of independence assumptions in the one-way ANOVA, *Am. Stat.*, **41**(2), 123–129.

Southam D. C. and Lewis J. E., (2013), Supporting alternative strategies for learning chemical applications of group theory, *J. Chem. Educ.*, **90**(11), 1425–1432.

Stevens J. P., (2007), One-way analysis of variance, in *Intermediate statistics: A modern approach*, New York, NY: Routledge, pp. 45–104.

Stitzel S. and Raje S., (2021), Understanding diverse needs and access to resources for student success in an introductory college chemistry course, *J. Chem. Educ.*, **99**(1), 49–55.

Theobald E., (2018), Students are rarely independent: when, why, and how to use random effects in discipline-based education research, *CBE—Life Sci. Educ.*, **17**(3), rm2.

Vitale M. P., Porzio G. C. and Doreian P., (2016), Examining the effect of social influence on student performance through network autocorrelation models, *J. Appl. Stat.*, **43**(1), 115–127.

Vyhnal C. R., Mahoney E. H., Lin Y., Radpour R. and Wadsworth H., (2020), Pigment synthesis and analysis of color in art: an example of applied science for high school and college chemistry students, *J. Chem. Educ.*, **97**(5), 1272–1282.

Wasserman S. and Faust K., (1994), Social network analysis in the social and behavioral sciences. in *Social network analysis: Methods and applications*, New York, NY: Cambridge University Press, pp. 3–27.

Webster J. M., Duvall J., Gaines L. M. and Smith R. H., (2003), The roles of praise and social comparison information in the experience of pride, *J. Soc. Psychol.*, **143**(2), 209–232.

Wellhöfer L. and Lühken A., (2022), Problem-based learning in an introductory inorganic laboratory: identifying connections between learner motivation and implementation, *J. Chem. Educ.*, **99**(2), 864–873.

Wesemann J., (2005), Undergraduate transitions: enhancing student success, *J. Chem. Educ.*, **82**(2) 196–198.

White R. W., (1959), Motivation reconsidered: the concept of competence, *Psychol. Rev.*, **66**(5), 297–333.

Wigfield A. and Cambria J., (2010), Achievement motivation, in Weiner, I. B. and Craighead, W. E., (ed.), *The Corsini encyclopedia of psychology*, Hoboken, NJ: John Wiley & Sons, Inc., pp. 14–15.

Williams G. C. and Deci E. L., (1996), Internalization of biopsychosocial values by medical students: a test of self-determination theory, *J. Pers. Soc. Psychol.*, **70**(4), 767.

Williams U. J. and Dries D. R., (2022), Supporting Fledgling Scientists: The Importance of Autonomy in a Guided-Inquiry Laboratory Course, *J. Chem. Educ.*, **99**(2), 701–707.

Wu F. and Teets T. S., (2021), Effects of the Covid-19 pandemic on student engagement in a general chemistry course, *J. Chem. Educ.*, **98**(12), 3633–3642.

Zhang C. and Conrad F., (2014), Speeding in web surveys: the tendency to answer very fast and its association with straightlining, *Surv. Res. Methods*, **8**(2), 127–135.

1024 | *Chem. Educ. Res. Pract.*, 2023, **24**, 1003–1024

This journal is © The Royal Society of Chemistry 2023